http://algs4.cs.princeton.edu

4.3 Minimum Spanning Trees

- *introduction*
- *greedy algorithm*
- *edge-weighted graph API*
- *Kruskal's algorithm*
- *Prim's algorithm*
- *context*

Algorithms

Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

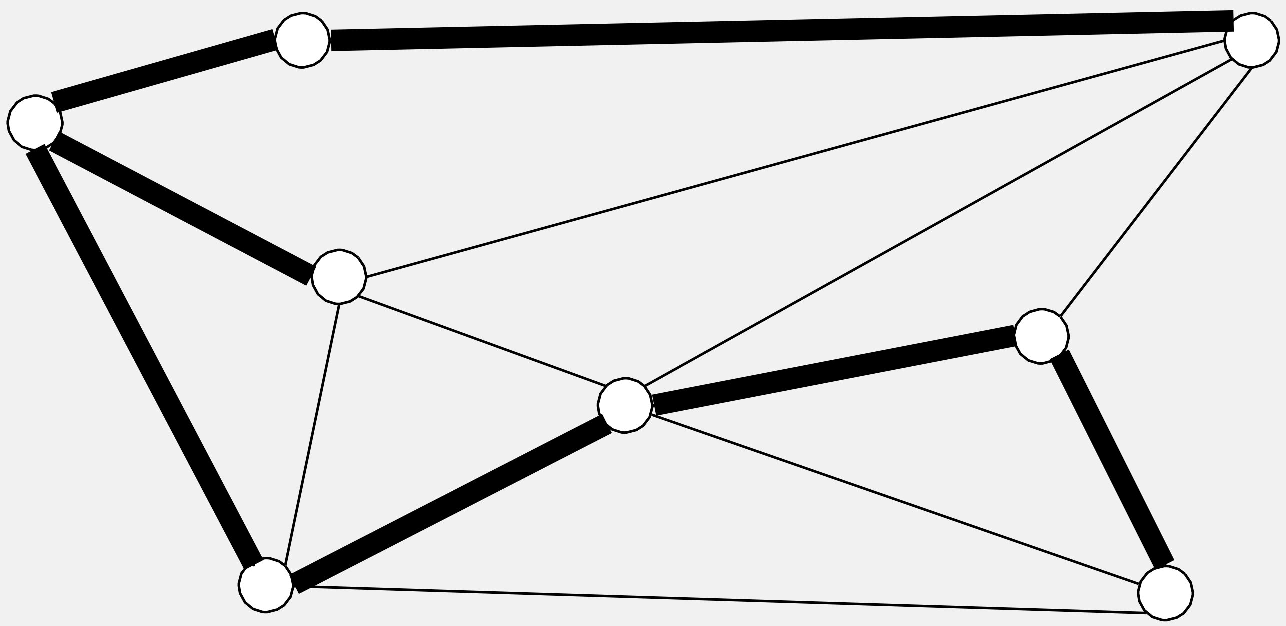
4.3 Minimum Spanning Trees

- *introduction*
- *greedy algorithm*
- *edge-weighted graph API*
- *Kruskal's algorithm*
- *Prim's algorithm*
- *context*

Minimum spanning tree

Definim. Një spanning tree i G është një nëngraf T i cili:

- është i lidhur.
- aciklik.
- përfshinë të gjitha kulmet.

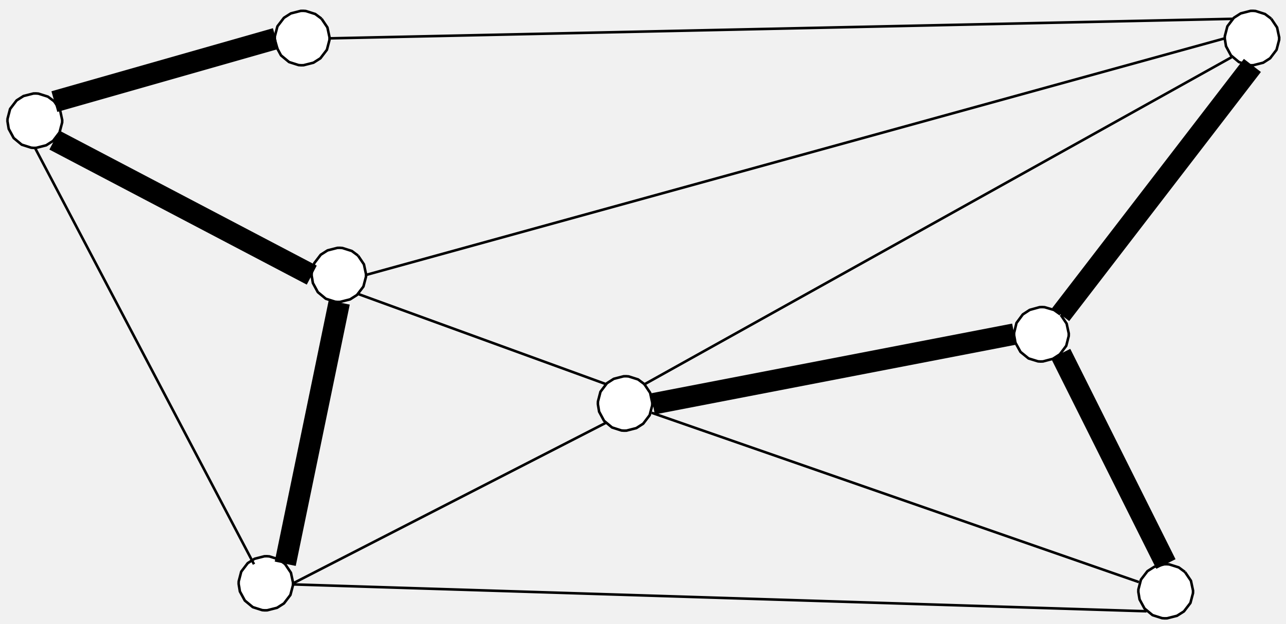


graph G

Minimum spanning tree

Definim. Një spanning tree i G është një nëngraf T i cili:

- është i lidhur.
- aciklik.
- përfshinë të gjitha kulmet.

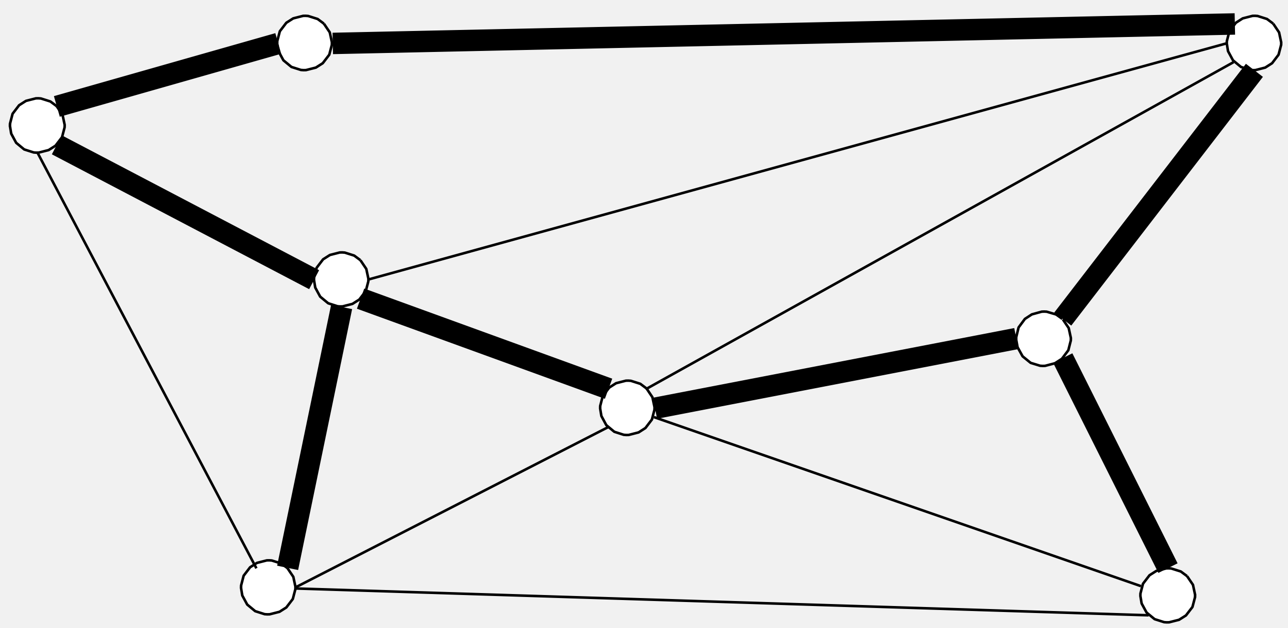


Nuk është lidhur

Minimum spanning tree

Definim. A spanning tree i G është një nëngraf T i cili :

- është i lidhur.
- aciklik.
- përfshinë të gjitha kulmet.

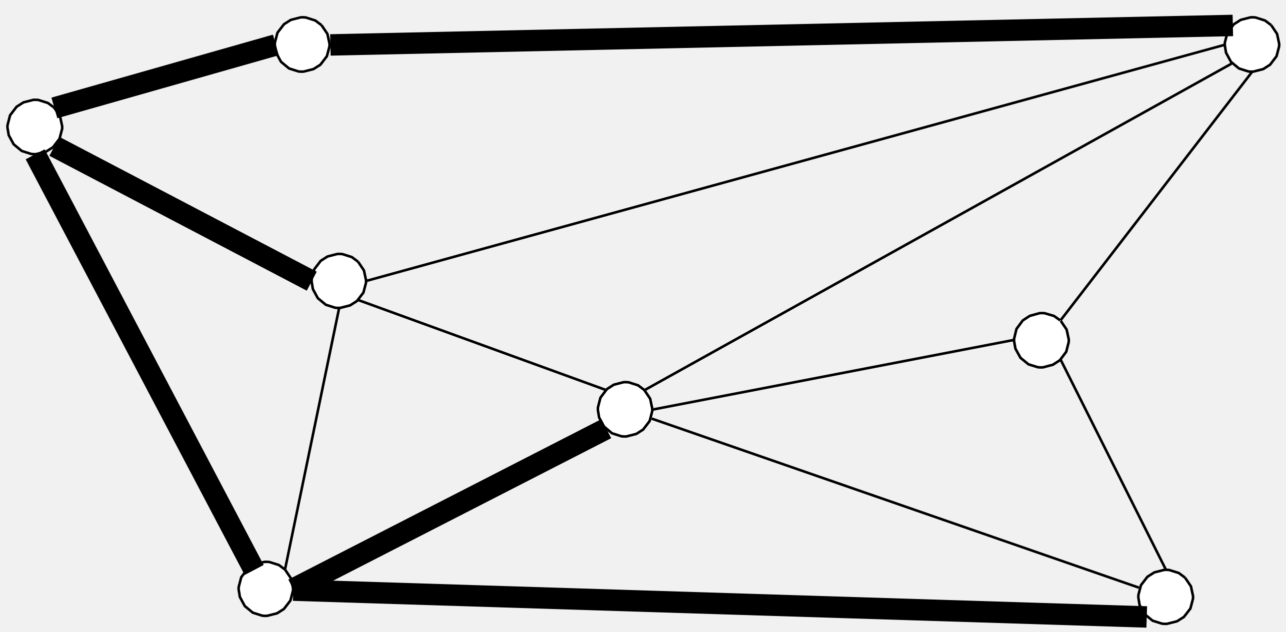


Nuk është aciklik

Minimum spanning tree

Definim. Një spanning tree i G është një nëngraf T i cili:

- është i lidhur.
- aciklik.
- përfshinë të gjitha kulmet.

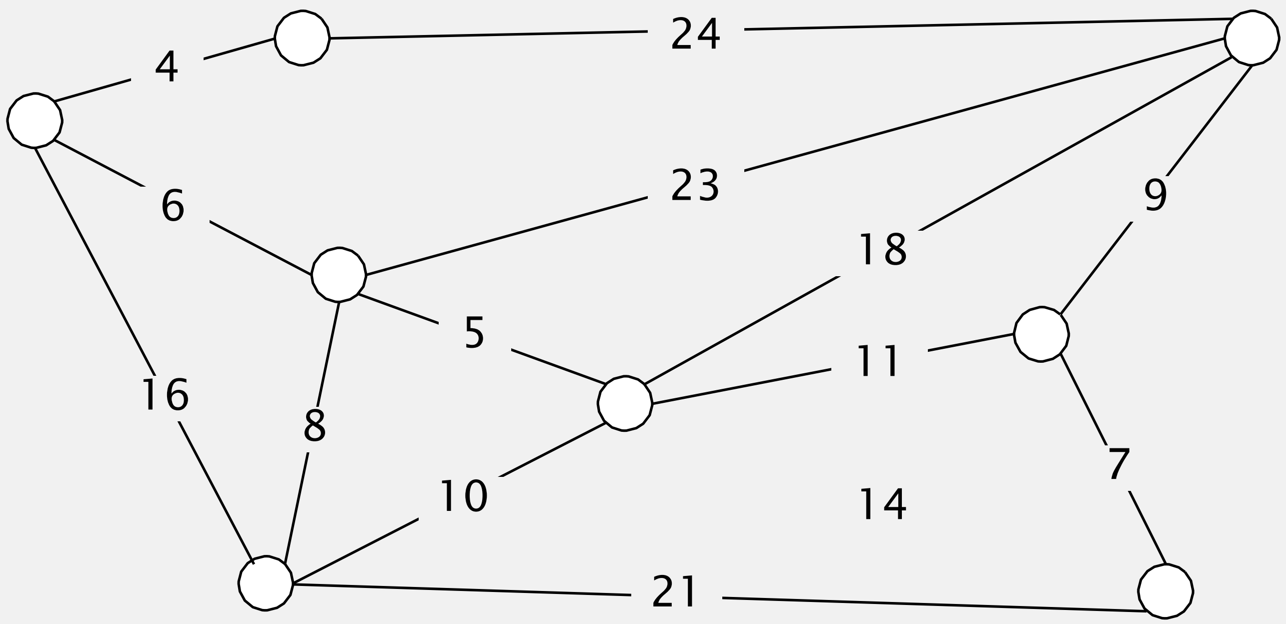


Nuk është spanning

Minimum spanning tree

E dhënë. Undirected graph G me vlera pozitive të segmenteve (të lidhura).

Qëllimi. Të gjindet një min weight spanning tree.

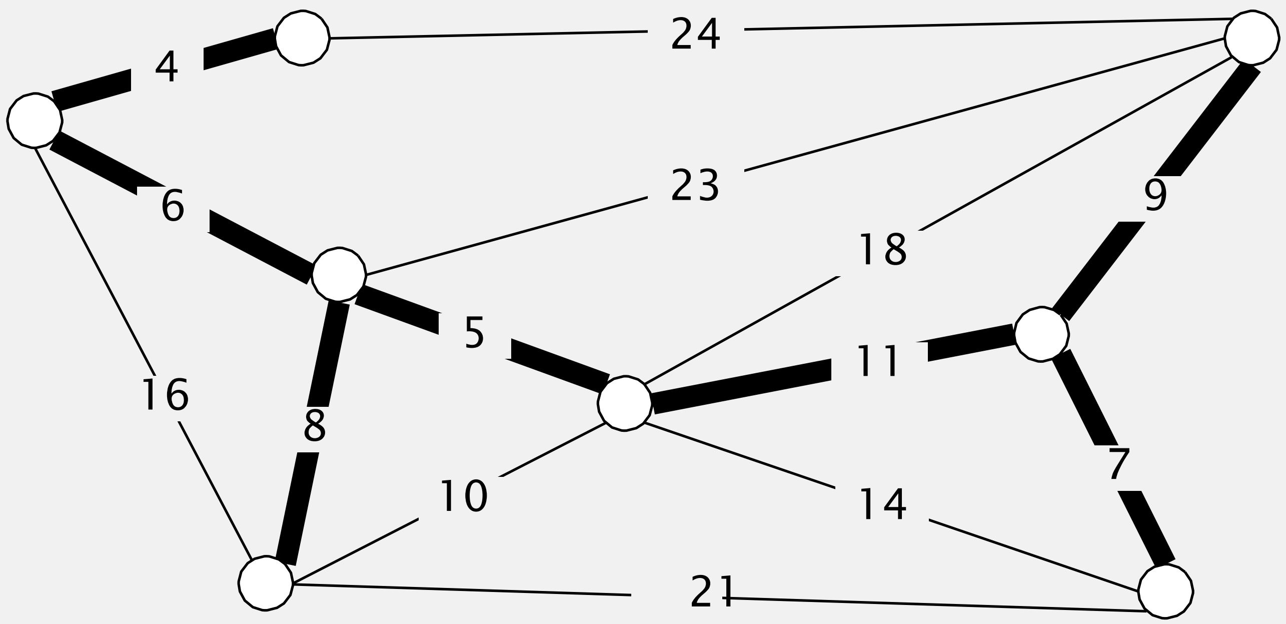


edge-weighted graph G

Minimum spanning tree

E dhënë. Undirected graph G me vlera pozitive të segmenteve (të lidhura).

Qëllimi. Të gjindet një min weight spanning tree.

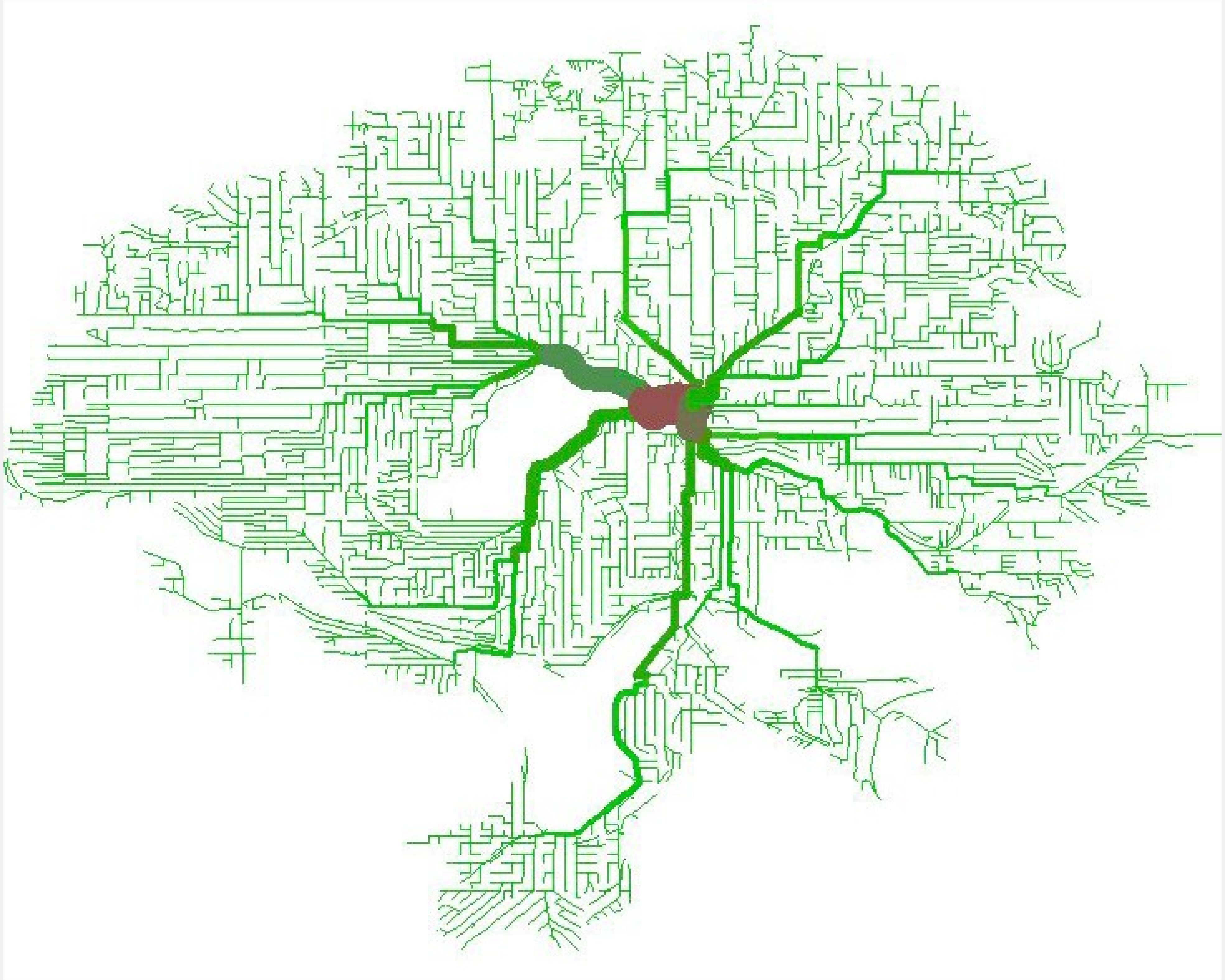


minimum spanning tree T
(çmimi = 50 = 4 + 6 + 8 + 5 + 11 + 9 + 7)

Brute force. Provo të gjitha spanning trees?

Network design

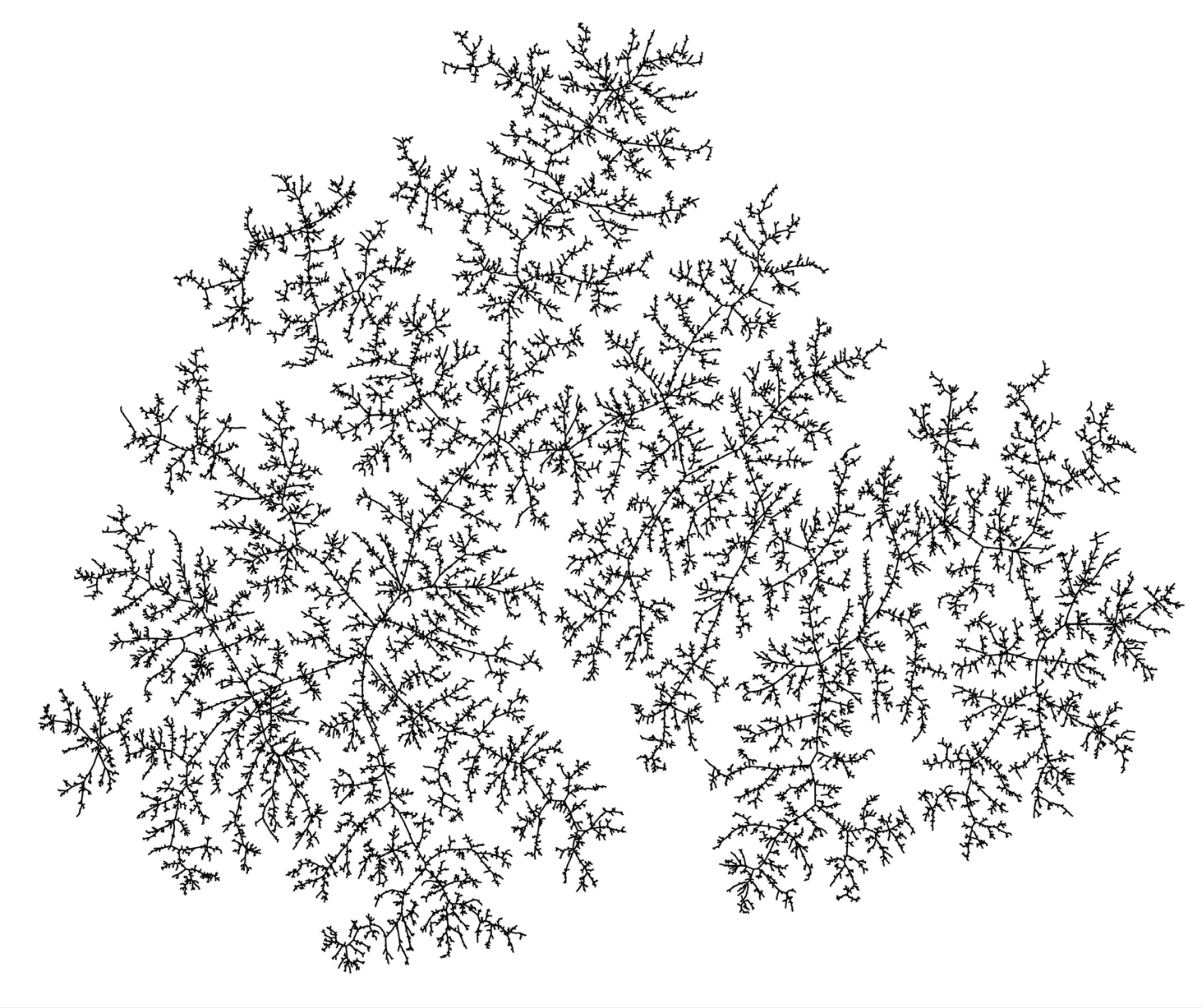
MST of bicycle routes in North Seattle



http://www.flickr.com/photos/ewedistrict/21980840

Models of nature

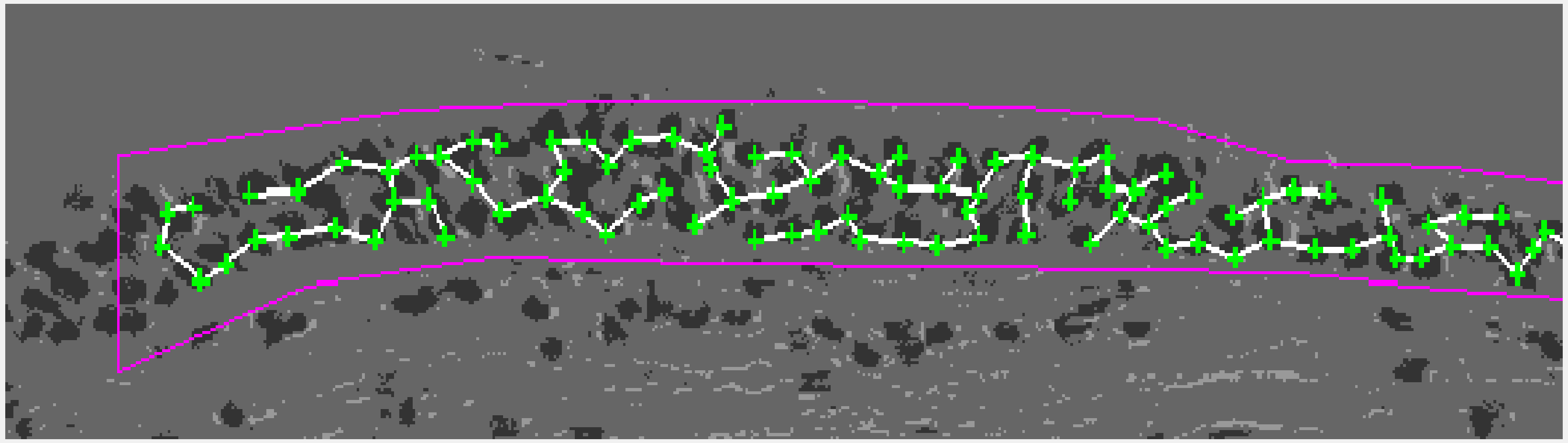
MST of random graph

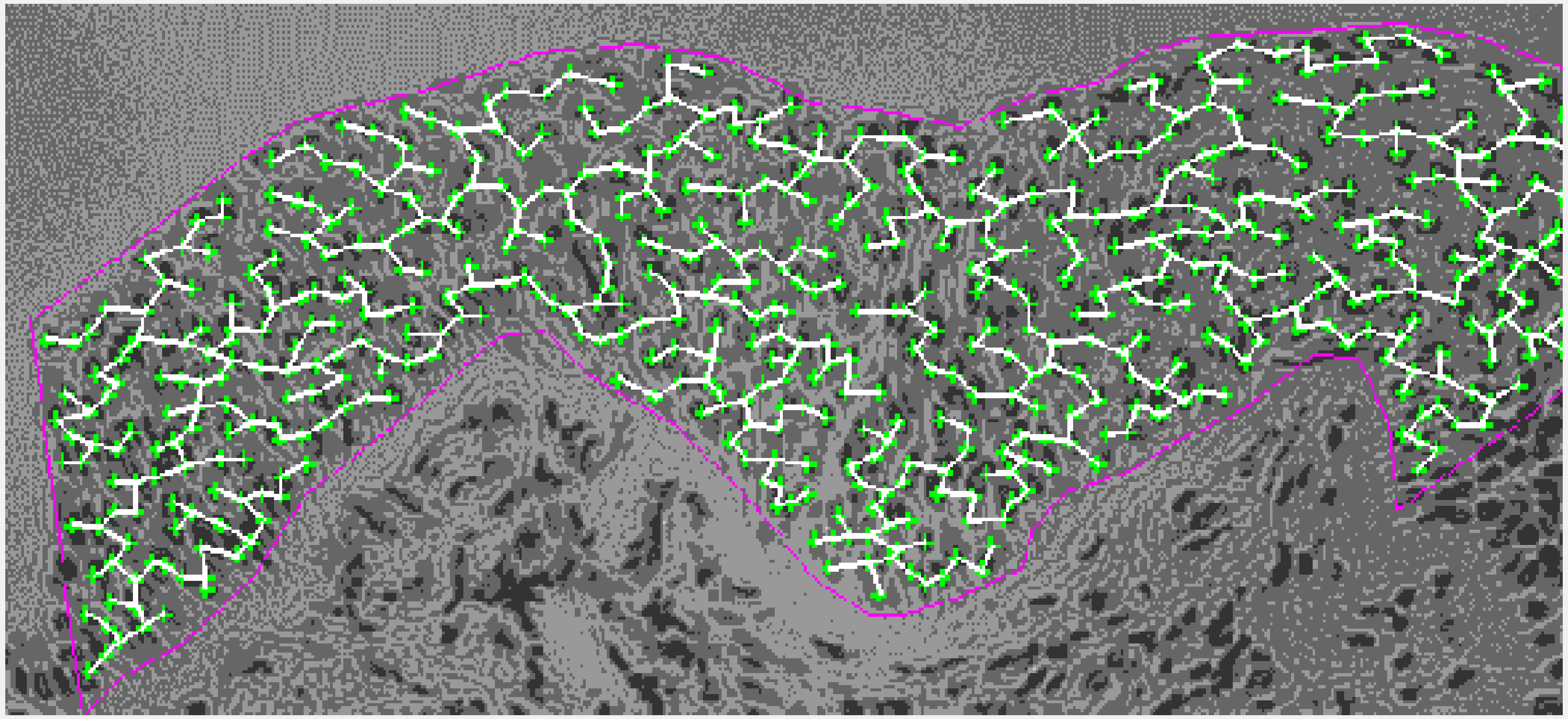


http://algo.inria.fr/broutin/gallery.html

Medical image processing

MST describes arrangement of nuclei in the epithelium for cancer research

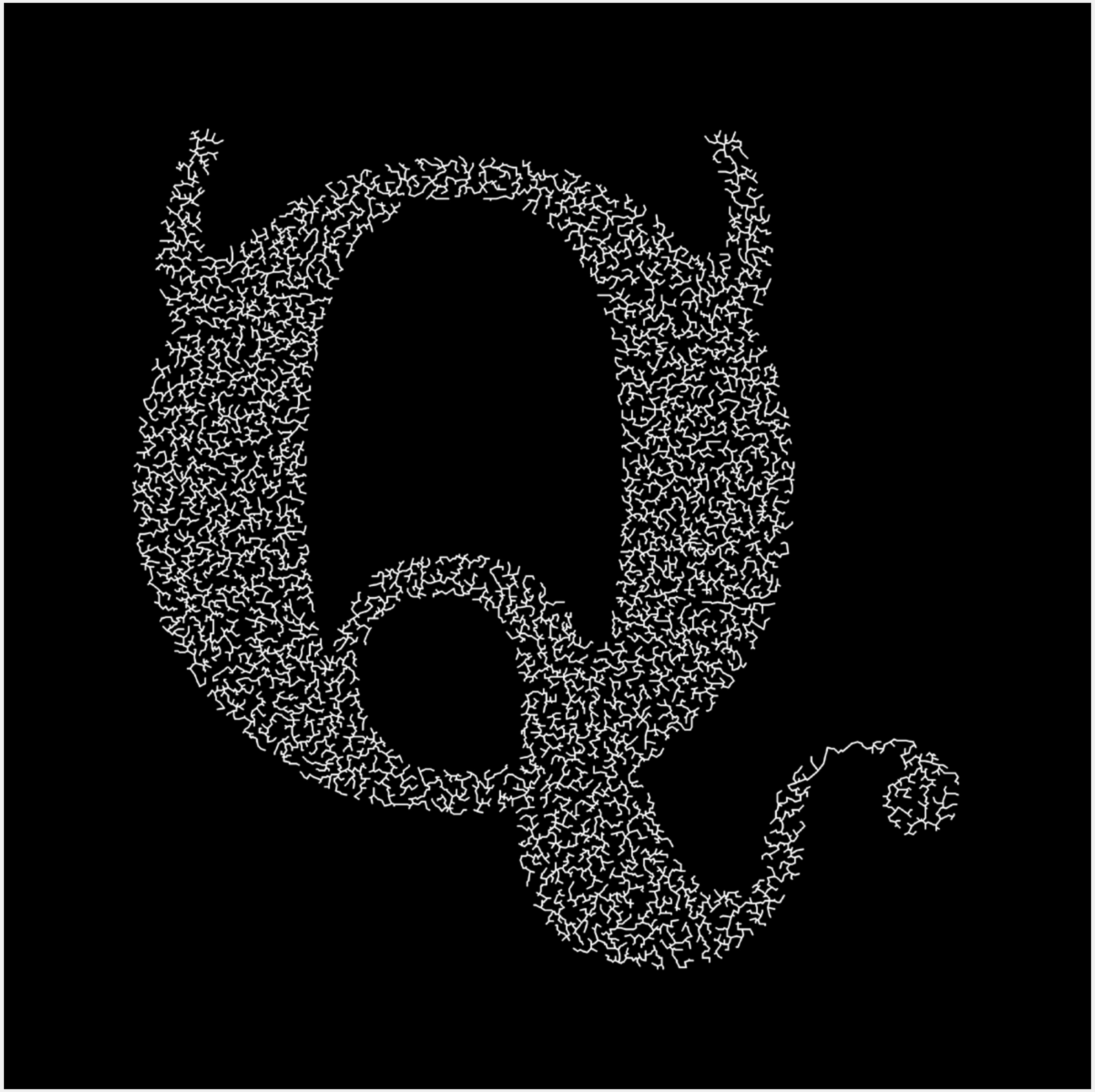




http://www.bccrc.ca/ci/ta01_archlevel.html

Medical image processing

MST dithering



http://www.flickr.com/photos/quasimondo/2695389651

Applications

MST is fundamental problem with diverse applications.

- Dithering.
- Cluster analysis.
- Max bottleneck paths.
- Real-time face verification.
- LDPC codes for error correction.
- Image registration with Renyi entropy.
- Find road networks in satellite and aerial imagery.
- Reducing data storage in sequencing amino acids in a protein.
- Model locality of particle interactions in turbulent fluid flows.
- Autoconfig protocol for Ethernet bridging to avoid cycles in a network.
- Approximation algorithms for NP-hard problems (e.g., TSP, Steiner tree).
- Network design (communication, electrical, hydraulic, computer, road).

http://www.ics.uci.edu/~eppstein/gina/mst.html

http://algs4.cs.princeton.edu

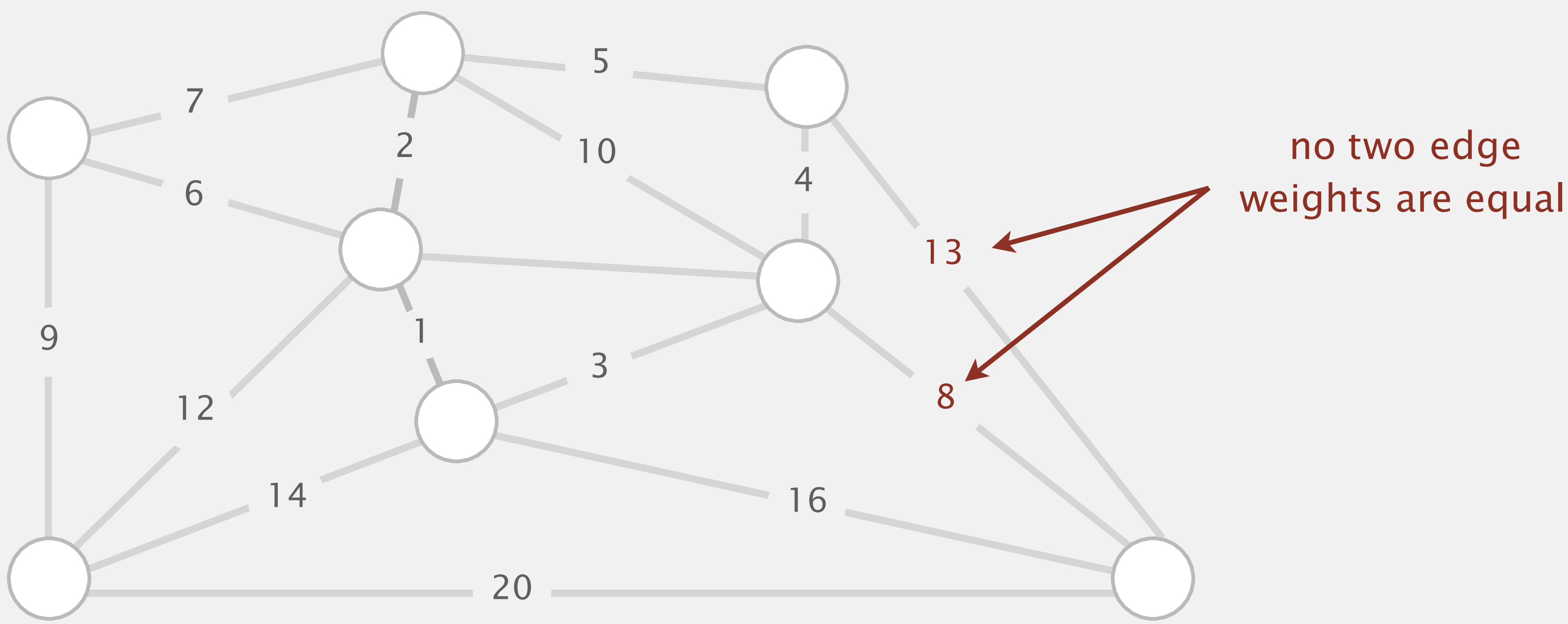
4.3 Minimum Spanning Trees

- *introduction*
- *greedy algorithm*
- *edge-weighted graph API*
- *Kruskal's algorithm*
- *Prim's algorithm*
- *context*

Supozim i thjeshtësuar

- Grafi është i lidhur.
- Pesha e segmenteve është e veçantë.

Rezultati. MST ekziston dhe është unik.

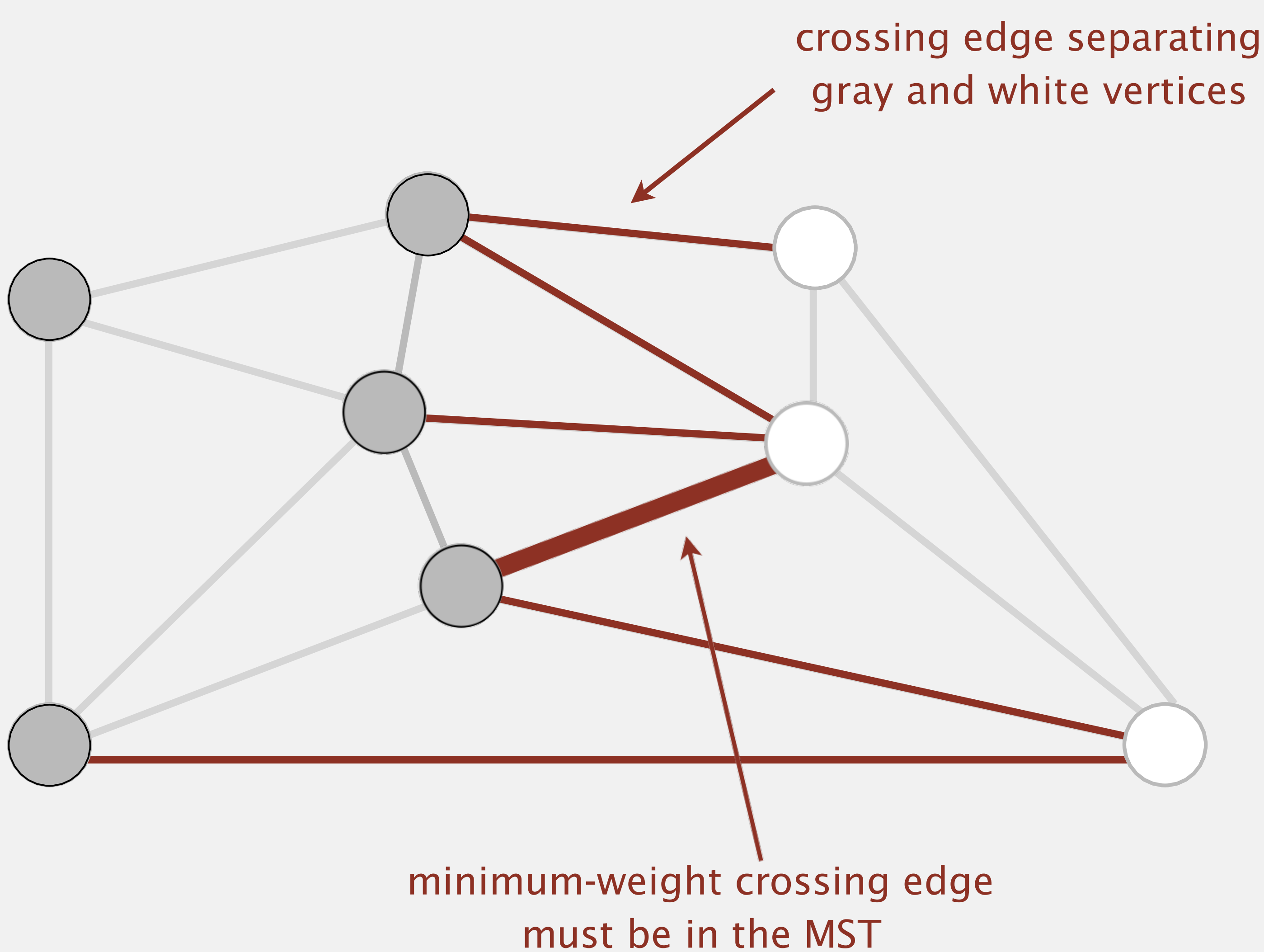


Cut - vetitë

Definim. Një cut në një graf është ndarja e kulmeve në dy bashkësi (jo boshe).

Definim. Një crossing edge lidhë një kulm në njërën bashkësi me një kulm në tjetrën.

Cut vetitë. Për çdo cut, crossing edge i min weight është në MST.



Cut – vetitë : correctness proof

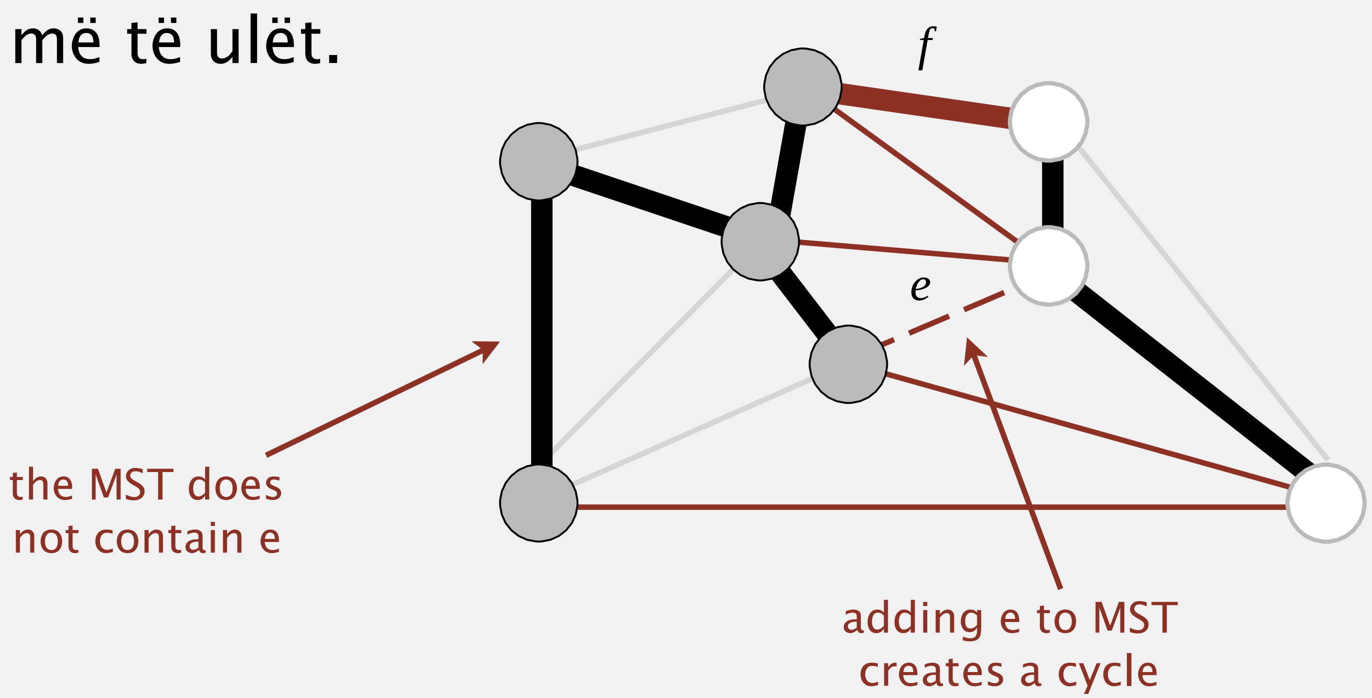
Definim. Një cut në një graf është ndarja e kulmeve në dy bashkësi (jo boshe).

Definim. Një crossing edge lidhë një kulm në njërën bashkësi me një kulm në tjetrën.

Cut - vetitë. Për çdo cut, crossing edge i min weight është në MST.

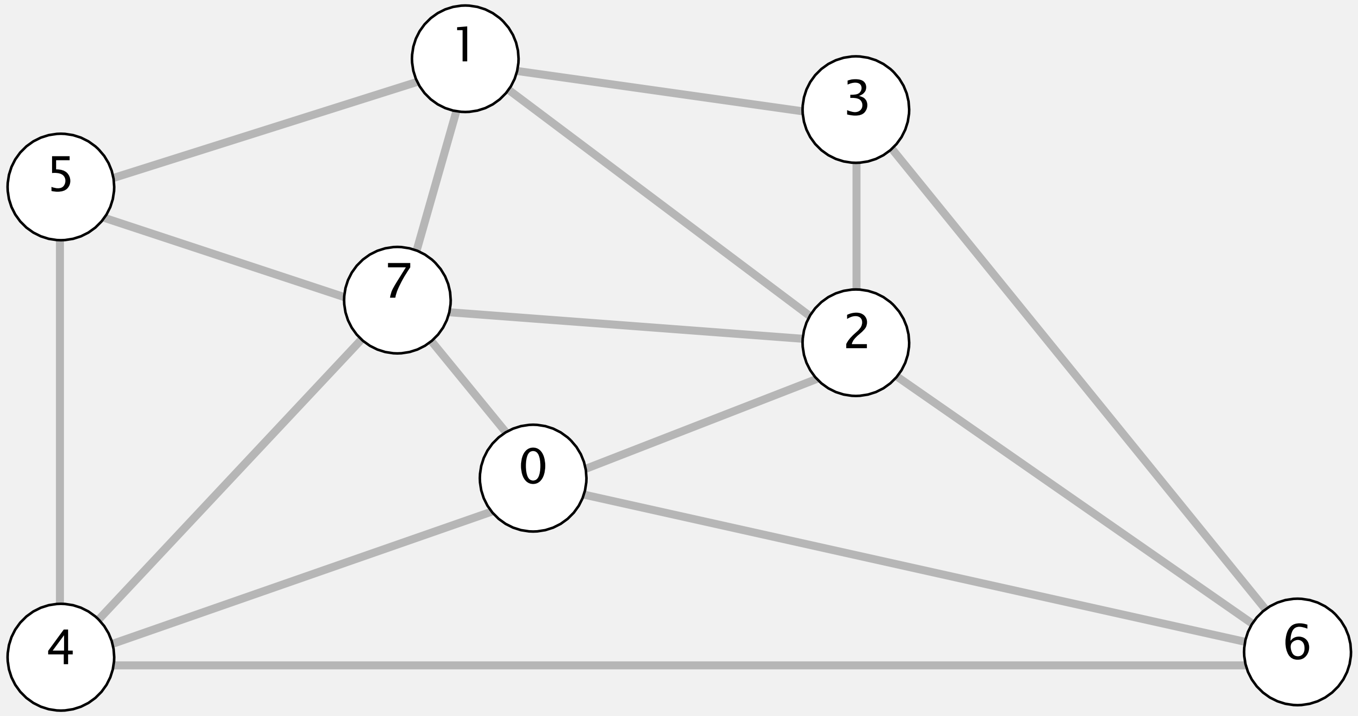
Vërtetim. Supuzo që min-weight crossing edge e nuk është në MST.

- Shtimi i e në MST krijon një cikël.
- Një tjetër segment f në cikël duhet të jetë një crossing edge.
- Fshirja e f dhe shtimi i e është gjithashtu një spanning tree.
- Pasi që pasha e e është më e vogël sesa ajo e f, ai spanning tree është me peshë më të ulët.
- Kontradiktë.



Greedy MST algorithm demo

- Fillo kur të gjitha segmented janë të përhimta.
- Gjej cut pa crossing edges të zeza; ngjyrose min-weight edge në të zezë.
- Përsërit derisa $V - 1$ segmente janë të ngjyrosura zi.

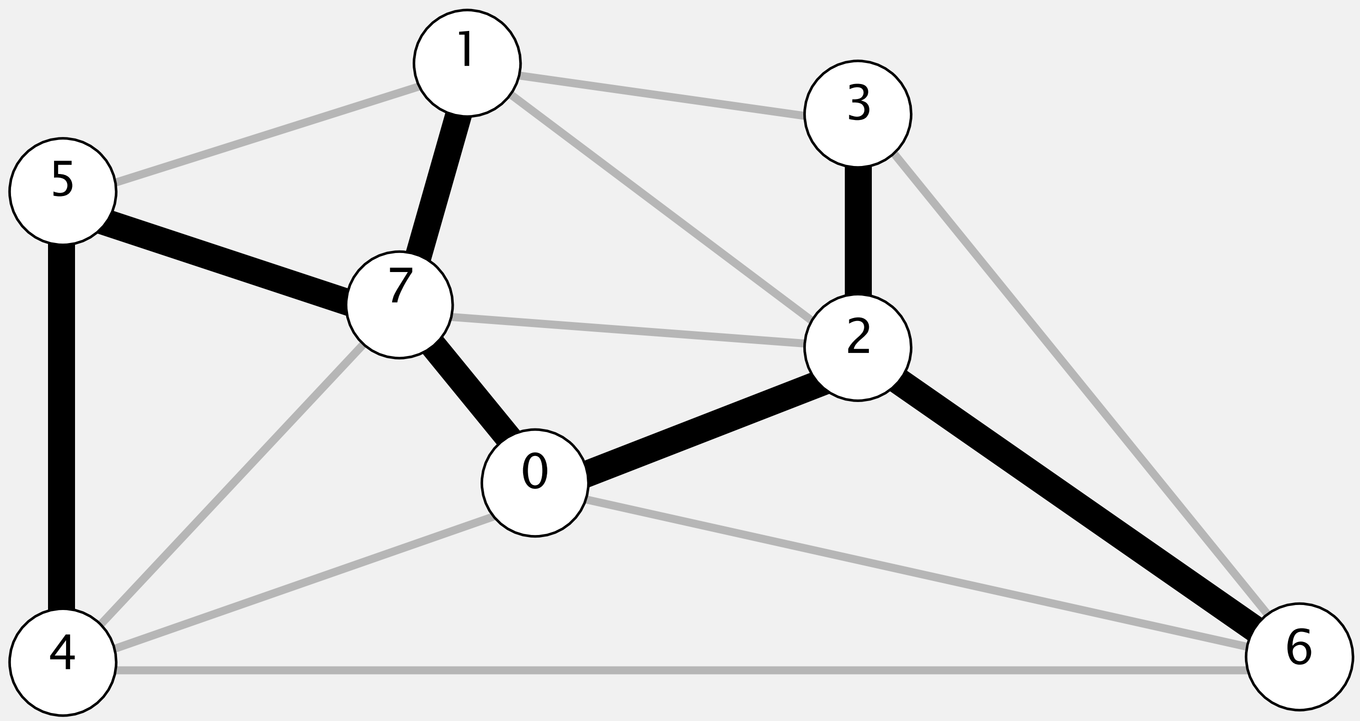


an edge-weighted graph

0-7	0.16
2-3	0.17
1-7	0.19
0-2	0.26
5-7	0.28
1-3	0.29
1-5	0.32
2-7	0.34
4-5	0.35
1-2	0.36
4-7	0.37
0-4	0.38
6-2	0.40
3-6	0.52
6-0	0.58
6-4	0.93

Greedy MST algorithm demo

- Fillo kur të gjitha segmented janë të përhimta.
- Gjej cut pa crossing edges të zeza; ngjyrose min-weight edge në të zezë.
- Përsërit derisa $V - 1$ segmente janë të ngjyrosura zi.



MST edges

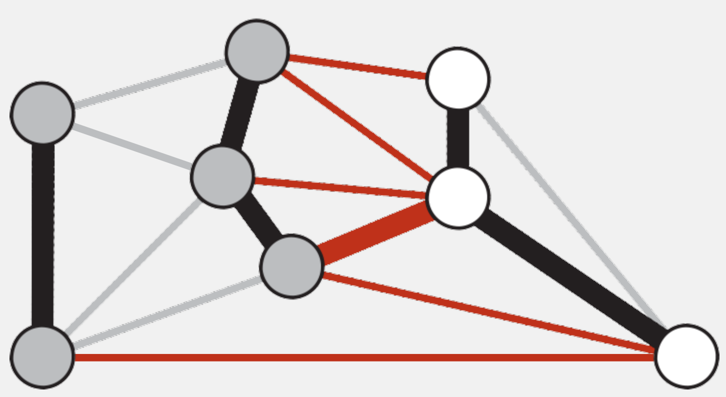
0–2 5–7 6–2 0–7 2–3 1–7 4–5

Greedy MST algorithm:

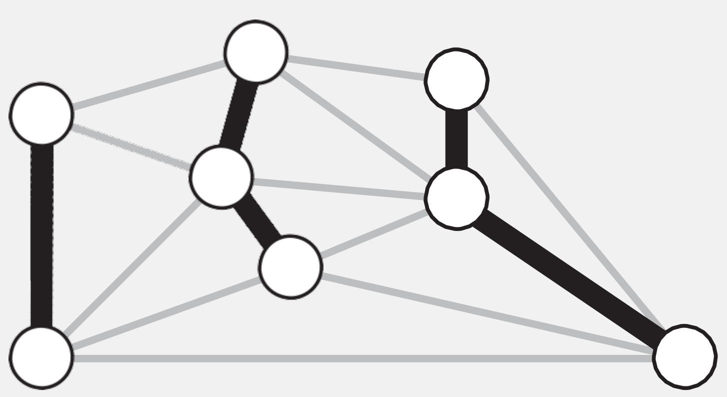
Teoremë. greedy algoritmi llogaritë MST.

Vërtetim.

- Cilido segment i zi është në MST (bazuar në vetitë e cut).
- Më pak se $V-1$ segmente të zeza $\Rightarrow$ cut pa crossing edges të zi.



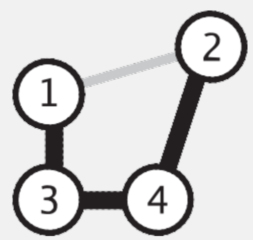
a cut with no black crossing edges



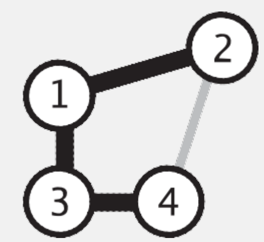
fewer than V-1 edges colored black

Removing two simplifying assumptions

Pyetje. Çka nëse pashat e segmenteve nuk dallojnë?
Përgjigje. Greedy MST algoritmi është i saktë nëse peshat janë të barabarta.

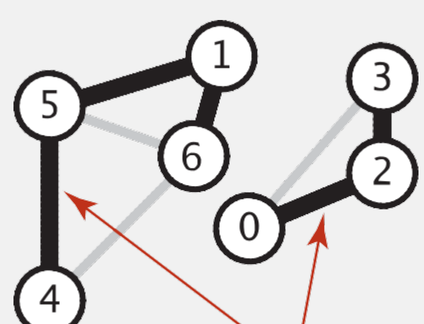


1	2	1.00
1	3	0.50
2	4	1.00
3	4	0.50



1	2	1.00
1	3	0.50
2	4	1.00
3	4	0.50

Q. Çka nëse grafi nuk është i lidhur?
A. Llogaritet minimum spanning forest = MST i secilit përbërës.



can independently compute MSTs of components

4	5	0.61
4	6	0.62
5	6	0.88
1	5	0.11
2	3	0.35
0	3	0.6
1	6	0.10
0	2	0.22

4.3 Minimum Spanning Trees

- introduction
- greedy algorithm
- *edge-weighted graph API*
- Kruskal's algorithm
- Prim's algorithm
- context

Algorithms

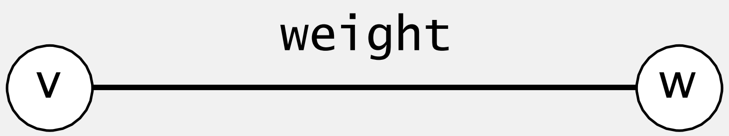
Robert Sedgewick | Kevin Wayne

http://algs4.cs.princeton.edu

Weighted edge API

Edge abstraction needed for weighted edges.

public class Edge implements Comparable<Edge>		
	Edge(int v, int w, double weight)	*create a weighted edge v-w*
int	either()	*either endpoint*
int	other(int v)	*the endpoint that's not v*
int	compareTo(Edge that)	*compare this edge to that edge*
double	weight()	*the weight*
String	toString()	*string representation*



Idiom for processing an edge `e`: `int v = e.either(), w = e.other(v);`

Weighted edge: Java implementation

```
public class Edge implements Comparable<Edge>
{
   private final int v, w;
   private final double weight;

   public Edge(int v, int w, double weight)
   {
      this.v = v;
      this.w = w;
      this.weight = weight;
   }

   public int either()
   {  return v;  }

   public int other(int vertex)
   {
      if (vertex == v) return w;
      else return v;
   }

   public int compareTo(Edge that)
   {
      if      (this.weight < that.weight) return -1;
      else if (this.weight > that.weight) return +1;
      else                                return  0;
   }
}
```

constructor

either endpoint

other endpoint

compare edges by weight

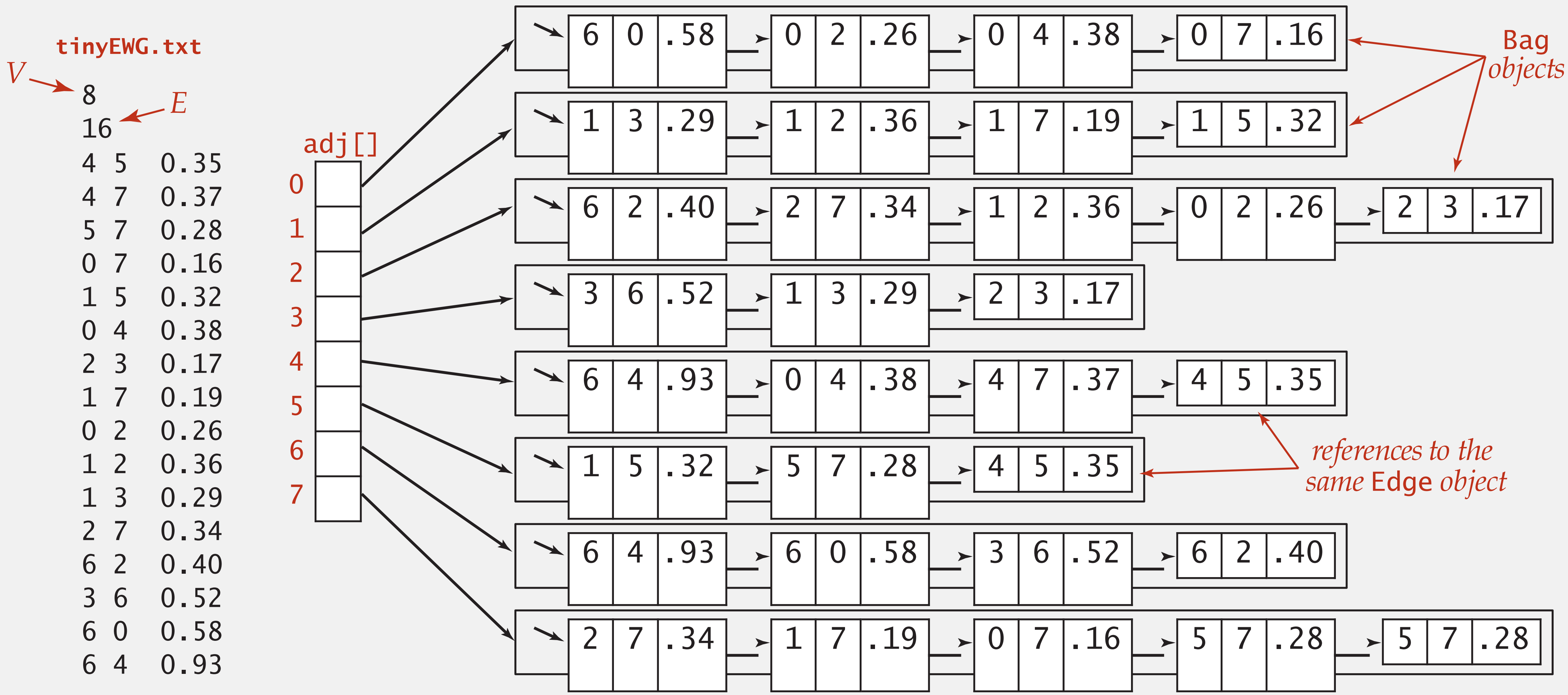
Edge-weighted graph API

public class	EdgeWeightedGraph	
	EdgeWeightedGraph(int V)	*create an empty graph with V vertices*
	EdgeWeightedGraph(In in)	*create a graph from input stream*
void	addEdge(Edge e)	*add weighted edge e to this graph*
Iterable<Edge>	adj(int v)	*edges incident to v*
Iterable<Edge>	edges()	*all edges in this graph*
int	V()	*number of vertices*
int	E()	*number of edges*
String	toString()	*string representation*

Conventions. Allow self-loops and parallel edges.

Edge-weighted graph: adjacency-lists representation

Maintain vertex-indexed array of Edge lists.



Edge-weighted graph: adjacency-lists implementation

```
public class EdgeWeightedGraph
{
   private final int V;
   private final Bag<Edge>[] adj;

   public EdgeWeightedGraph(int V)
   {
      this.V = V;
      adj = (Bag<Edge>[]) new Bag[V];
      for (int v = 0; v < V; v++)
         adj[v] = new Bag<Edge>();
   }

   public void addEdge(Edge e)
   {
      int v = e.either(), w = e.other(v);
      adj[v].add(e);
      adj[w].add(e);
   }

   public Iterable<Edge> adj(int v)
   {  return adj[v];  }
}
```

same as Graph, but adjacency lists of Edges instead of integers

constructor

add edge to both adjacency lists

Minimum spanning tree API

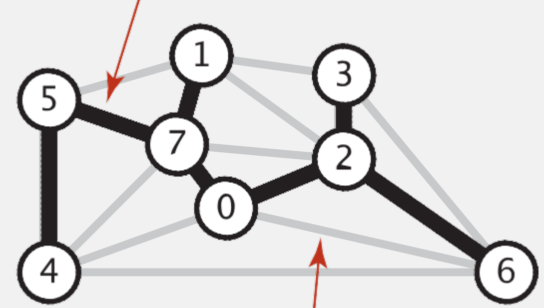
Q. How to represent the MST?

public class MST		
Iterable<Edge>	MST(EdgeWeightedGrap G)	*constructor*
	h edges()	*edges in MST*
double	weight()	*weight of MST*

tinyEWG.txt

```
V → 8
16 ← E
4 5  0.35
4 7  0.37
5 7  0.28
0 7  0.16
1 5  0.32
0 4  0.38
2 3  0.17
1 7  0.19
0 2  0.26
1 2  0.36
1 3  0.29
2 7  0.34
6 2  0.40
3 6  0.52
6 0  0.58
6 4  0.93
```

MST edge (black)



non-MST edge (gray)

```
% java MST tinyEWG.txt
0-7 0.16
1-7 0.19
0-2 0.26
2-3 0.17
5-7 0.28
4-5 0.35
6-2 0.40
1.81
```

Minimum spanning tree API

Q. How to represent the MST?

public class MST

Iterable<Edge>	MST(EdgeWeightedGrap G)	*constructor*
	h edges()	*edges in MST*
double	weight()	*weight of MST*

```
public static void main(String[] args)
{
   In in = new In(args[0]);
   EdgeWeightedGraph G = new EdgeWeightedGraph(in);
   MST mst = new MST(G);
   for (Edge e : mst.edges())
      StdOut.println(e);
   StdOut.printf("%.2f\n", mst.weight());
}
```

```
% java MST tinyEWG.txt
0-7 0.16
1-7 0.19
0-2 0.26
2-3 0.17
5-7 0.28
4-5 0.35
6-2 0.40
1.81
```

http://algs4.cs.princeton.edu

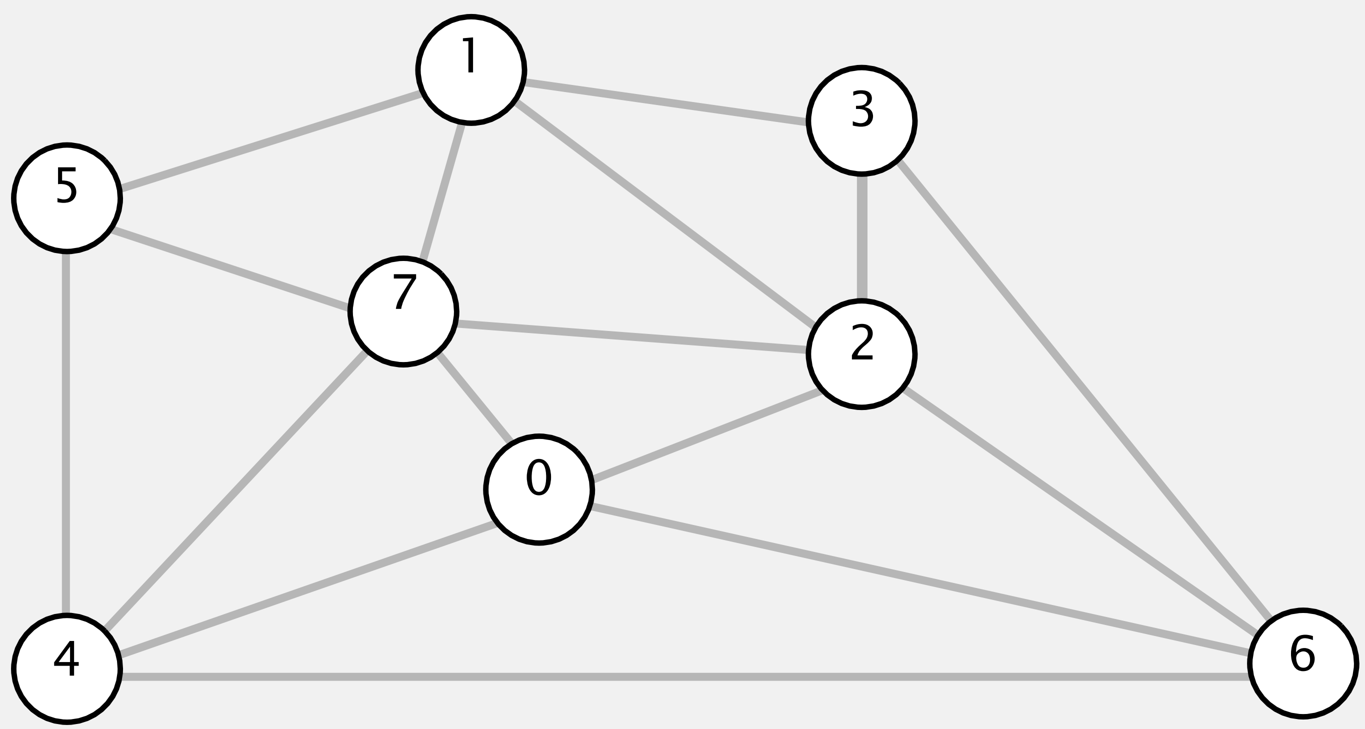
4.3 Minimum Spanning Trees

- introduction
- greedy algorithm
- edge-weighted graph API
- Kruskal's algorithm
- Prim's algorithm
- context

Kruskal's algorithm demo

Consider edges in ascending order of weight.

- Add next edge to tree T unless doing so would create a cycle.



an edge-weighted graph

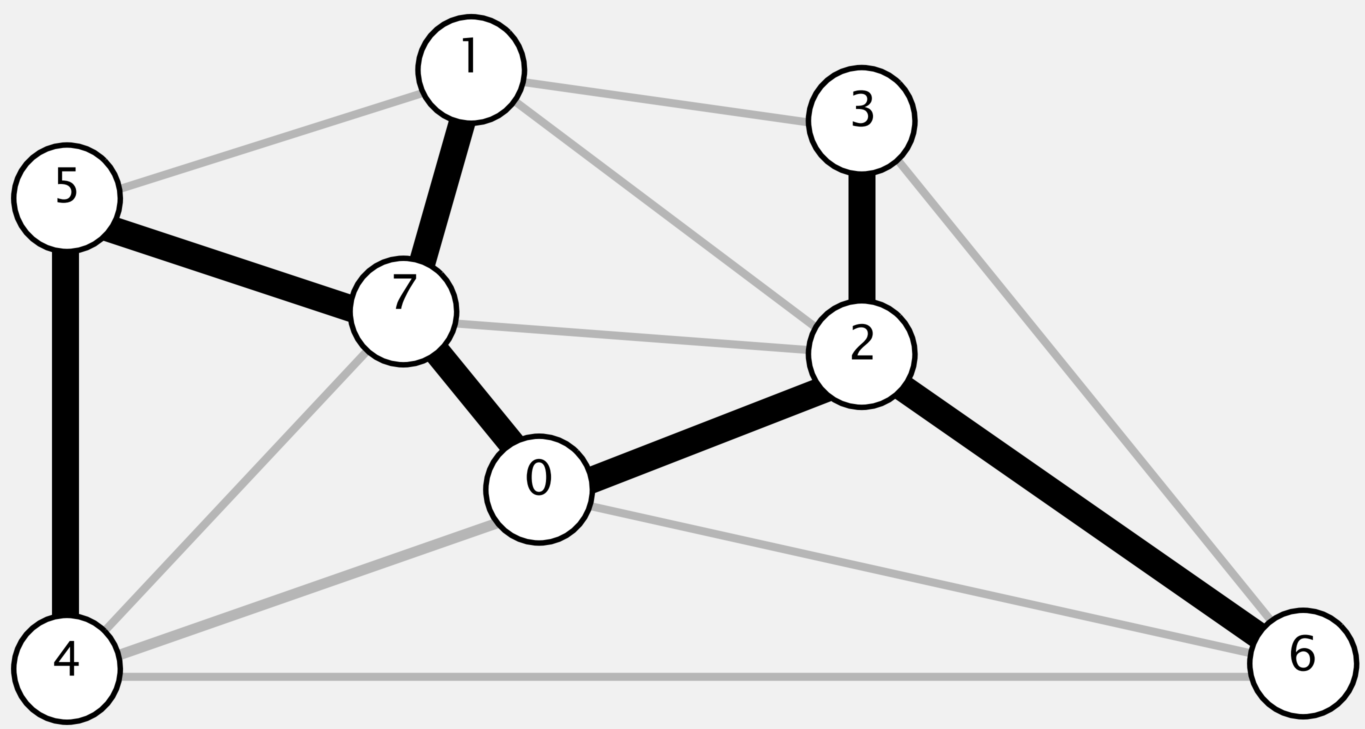
graph edges
sorted by weight

0-7	0.16
2-3	0.17
1-7	0.19
0-2	0.26
5-7	0.28
1-3	0.29
1-5	0.32
2-7	0.34
4-5	0.35
1-2	0.36
4-7	0.37
0-4	0.38
6-2	0.40
3-6	0.52
6-0	0.58
6-4	0.93

Kruskal's algorithm demo

Consider edges in ascending order of weight.

- Add next edge to tree T unless doing so would create a cycle.



a minimum spanning tree

0-7	0.16
2-3	0.17
1-7	0.19
0-2	0.26
5-7	0.28
1-3	0.29
1-5	0.32
2-7	0.34
4-5	0.35
1-2	0.36
4-7	0.37
0-4	0.38
6-2	0.40
3-6	0.52
6-0	0.58
6-4	0.93

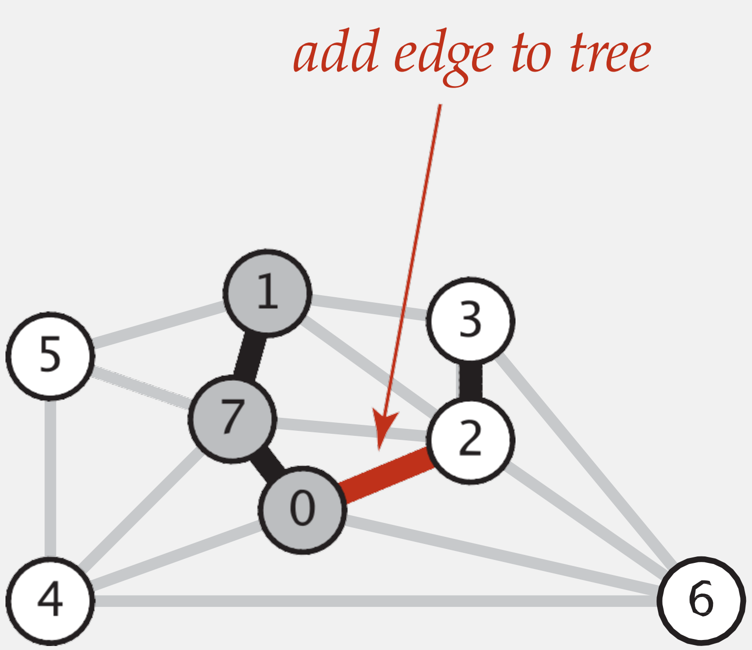
Kruskal's algorithm: visualization

Kruskal's algorithm: correctness proof

Proposition. [Kruskal 1956] Kruskal's algorithm computes the MST.

Pf. Kruskal's algorithm is a special case of the greedy MST algorithm.

- Suppose Kruskal's algorithm colors the edge $e = v\text{–}w$ black.
- Cut = set of vertices connected to v in tree T.
- No crossing edge is black.
- No crossing edge has lower weight. Why?

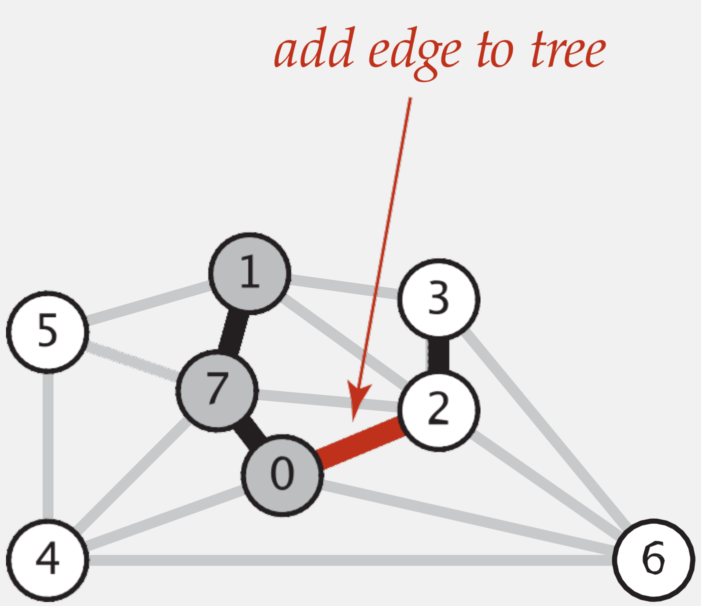


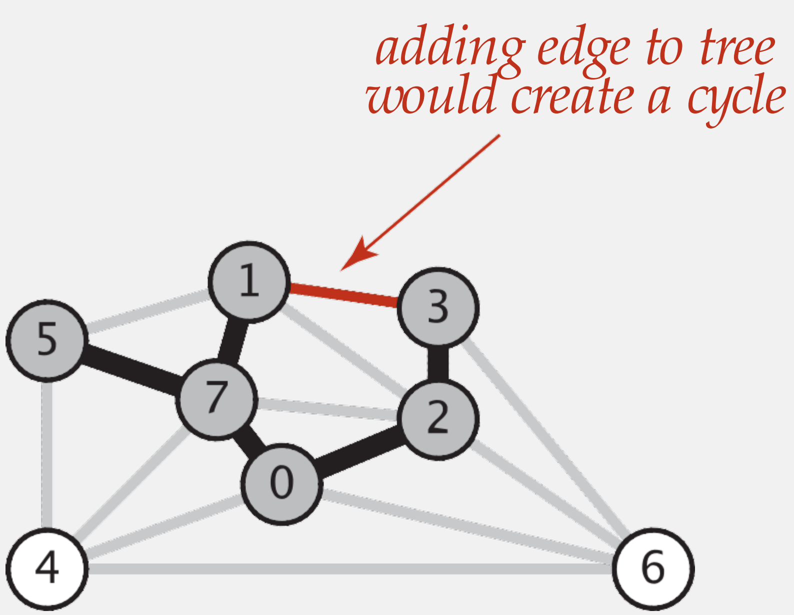
Kruskal's algorithm: implementation challenge

Challenge. Would adding edge $v–w$ to tree T create a cycle? If not, add it.

How difficult?

- $E + V$
- V ← run DFS from v, check if w is reachable (T has at most V – 1 edges)
- $\log V$
- $\log^* V$ ← use the union-find data structure !
- 1



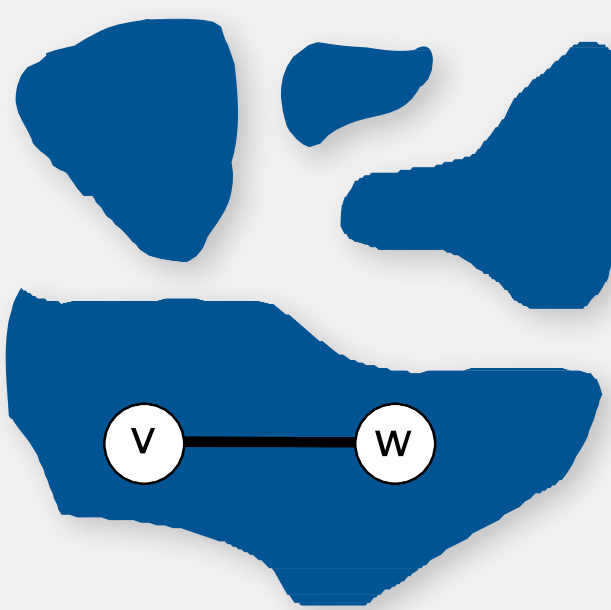


Kruskal's algorithm: implementation challenge

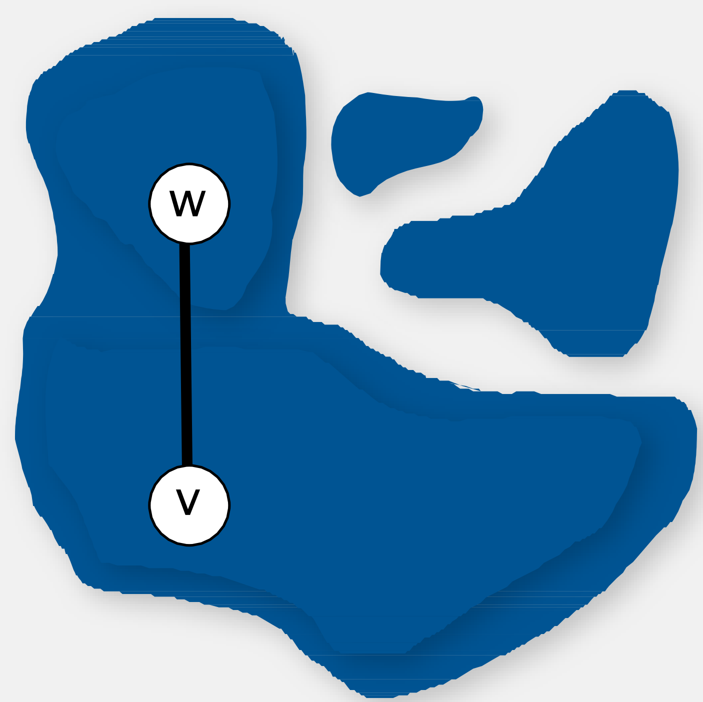
Challenge. Would adding edge $v–w$ to tree T create a cycle? If not, add it.

Efficient solution. Use the union-find data structure.

- Maintain a set for each connected component in T.
- If v and w are in same set, then adding $v–w$ would create a cycle.
- To add $v–w$ to T, merge sets containing v and w.



Case 1: adding v–w creates a cycle



Case 2: add v–w to T and merge sets containing v and w

Kruskal's algorithm: Java implementation

```
public class KruskalMST
{
   private Queue<Edge> mst = new Queue<Edge>();

   public KruskalMST(EdgeWeightedGraph G)
   {
      MinPQ<Edge> pq = new MinPQ<Edge>(G.edges());

      UF uf = new UF(G.V());
      while (!pq.isEmpty() && mst.size() < G.V()-1)
      {
         Edge e = pq.delMin();
         int v = e.either(), w = e.other(v);
         if (!uf.connected(v, w))
         {
            uf.union(v, w);
            mst.enqueue(e);
         }
      }
   }

   public Iterable<Edge> edges()
   {  return mst;  }
}
```

build priority queue (or sort)

greedily add edges to MST

edge v-w does not create cycle

merge sets

add edge to MST

Kruskal's algorithm: running time

Proposition. Kruskal's algorithm computes MST in time proportional to $E \log E$ (in the worst case).

Pf.

operation	frequency	time per op
build pq	1	E
delete-min	E	$\log E$
union	V	$\log^* V$ †
connected	E	$\log^* V$ †

† amortized bound using weighted quick union with path compression

recall: $\log^* V \leq 5$ in this universe

Remark. If edges are already sorted, order of growth is $E \log^* V$.

Algorithms

ROBERT SEDGEWICK | KEVIN WAYNE

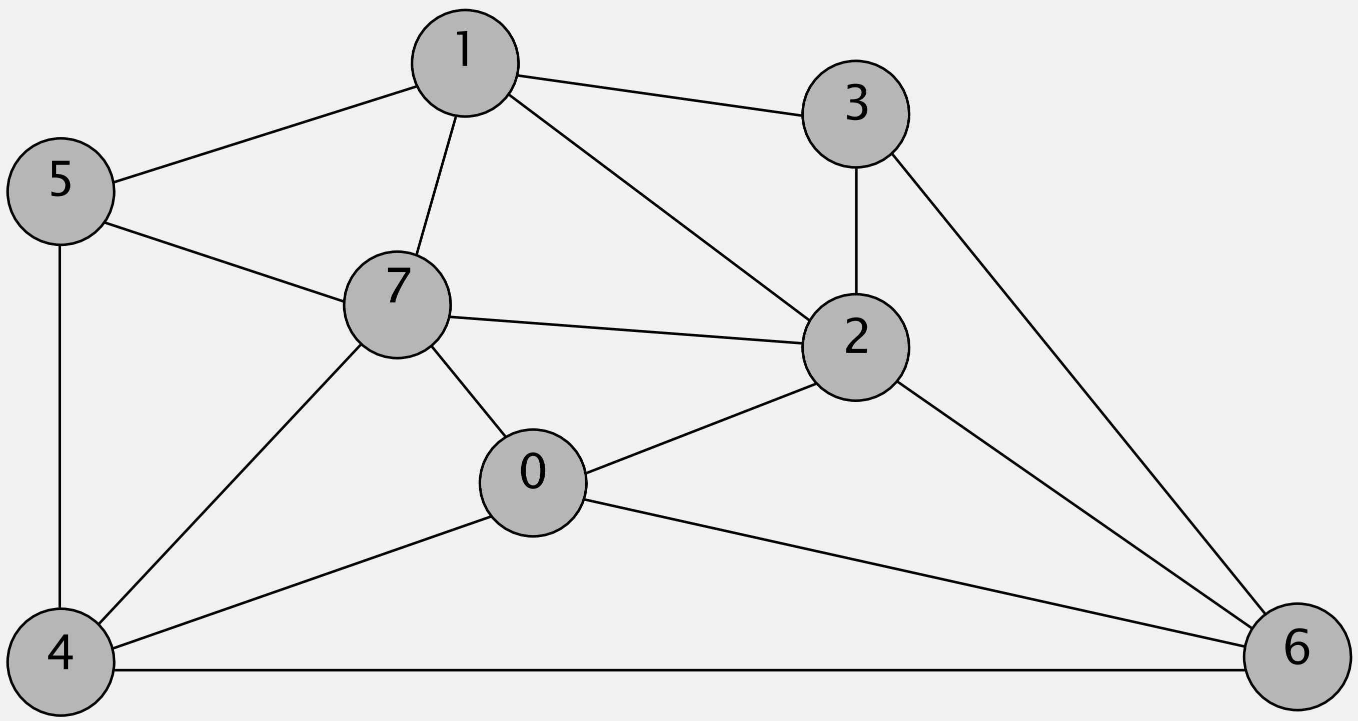
http://algs4.cs.princeton.edu

4.3 MINIMUM SPANNING TREES

- introduction
- greedy algorithm
- edge-weighted graph API
- Kruskal's algorithm
- *Prim's algorithm*
- context

Prim's algorithm demo

- Start with vertex 0 and greedily grow tree T.
- Add to T the min weight edge with exactly one endpoint in T.
- Repeat until $V - 1$ edges.

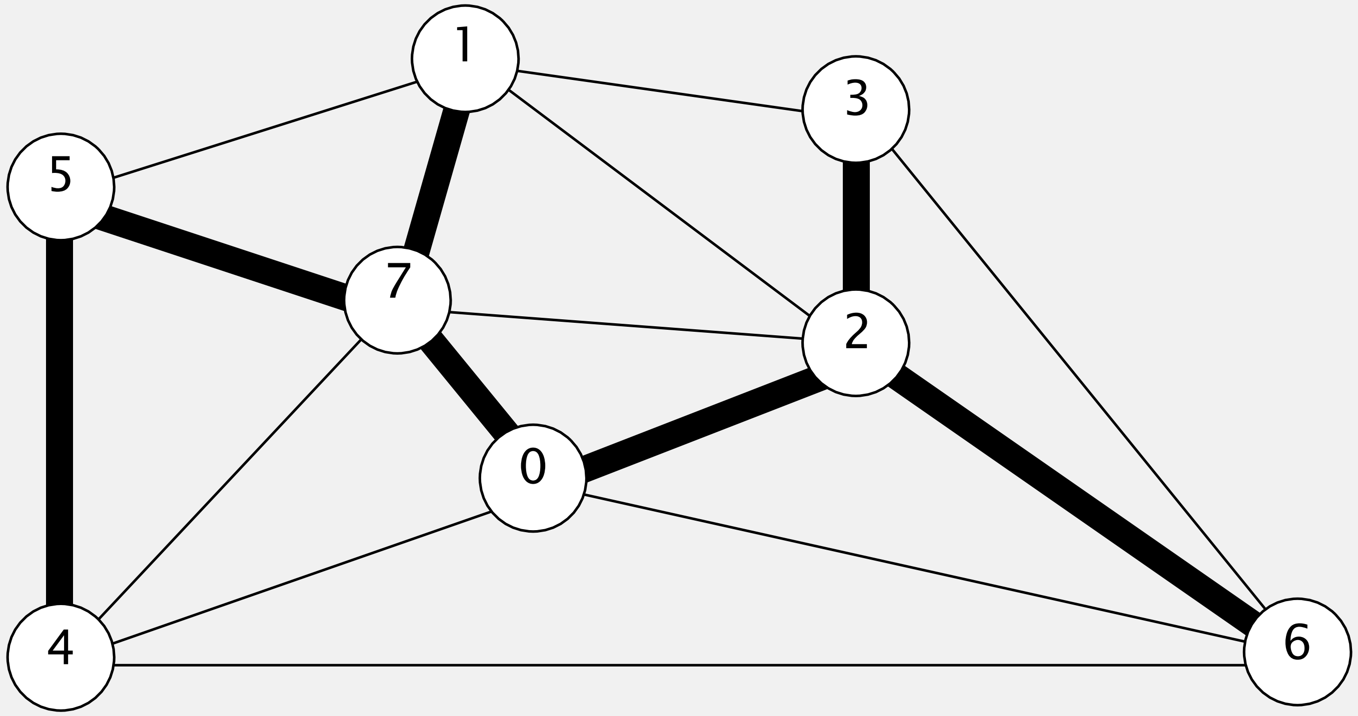


an edge-weighted graph

0-7	0.16
2-3	0.17
1-7	0.19
0-2	0.26
5-7	0.28
1-3	0.29
1-5	0.32
2-7	0.34
4-5	0.35
1-2	0.36
4-7	0.37
0-4	0.38
6-2	0.40
3-6	0.52
6-0	0.58
6-4	0.93

Prim's algorithm demo

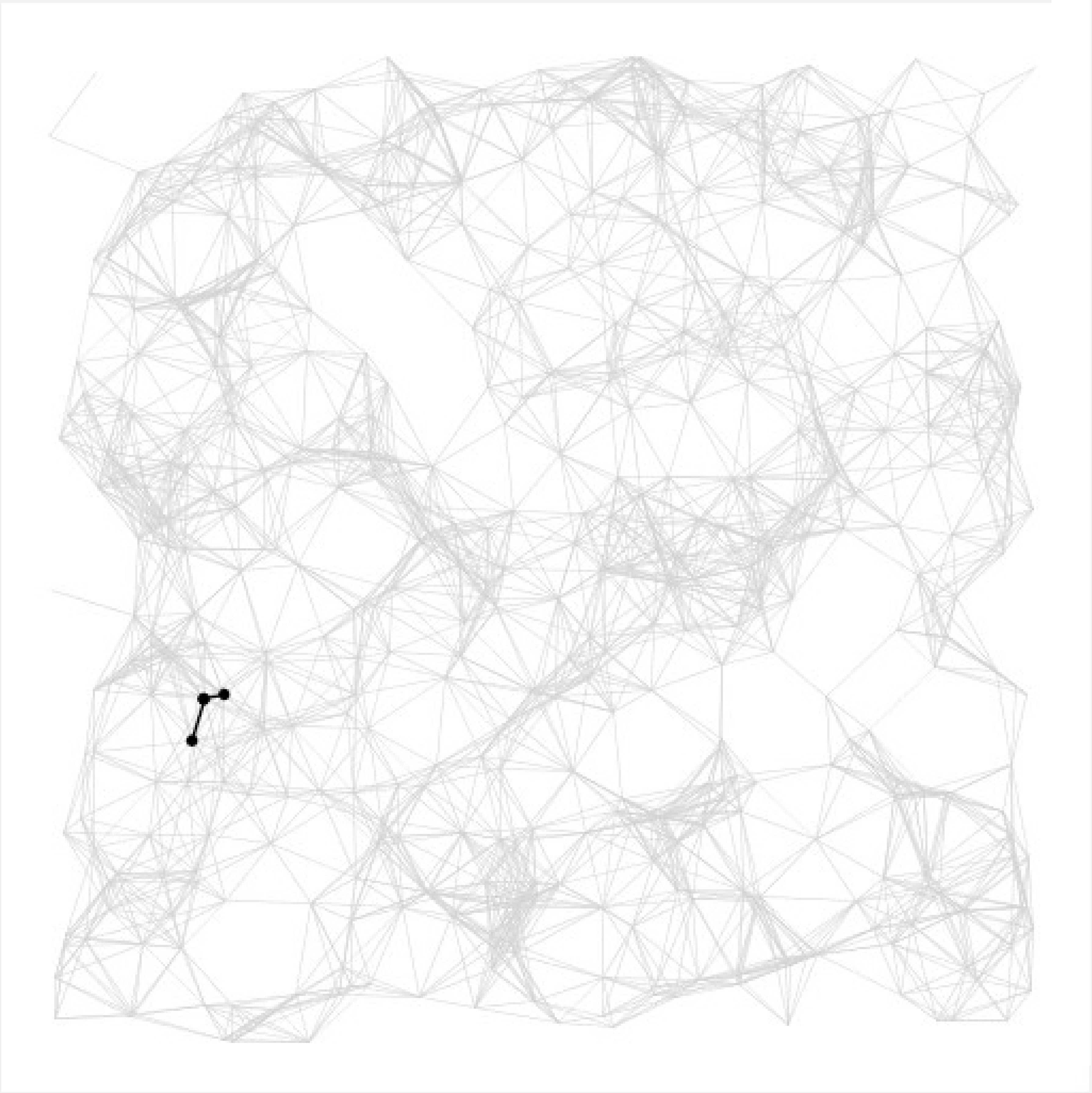
- Start with vertex 0 and greedily grow tree T.
- Add to T the min weight edge with exactly one endpoint in T.
- Repeat until $V - 1$ edges.



MST edges

0-7 1-7 0-2 2-3 5-7 4-5 6-2

Prim's algorithm: visualization

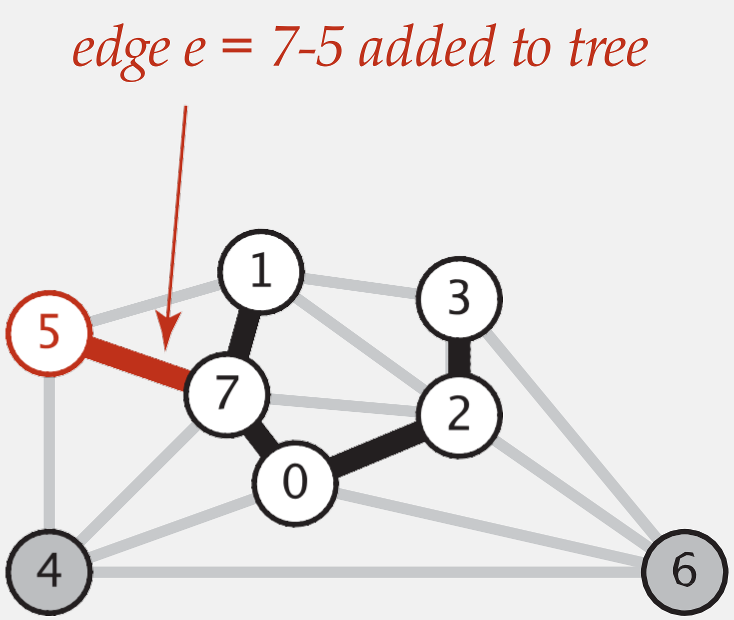


Prim's algorithm: proof of correctness

Proposition. [Jarník 1930, Dijkstra 1957, Prim 1959]
Prim's algorithm computes the MST.

Pf. Prim's algorithm is a special case of the greedy MST algorithm.

- Suppose edge e = min weight edge connecting a vertex on the tree to a vertex not on the tree.
- Cut = set of vertices connected on tree.
- No crossing edge is black.
- No crossing edge has lower weight.



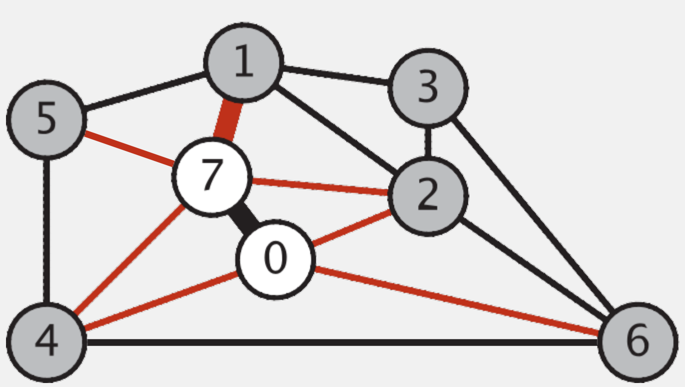
Prim's algorithm: implementation challenge

Challenge. Find the min weight edge with exactly one endpoint in T.

How difficult?

- E ← try all edges
- V
- $\log E$ ← use a priority queue!
- $\log^* E$
- 1

1-7 is min weight edge with exactly one endpoint in T



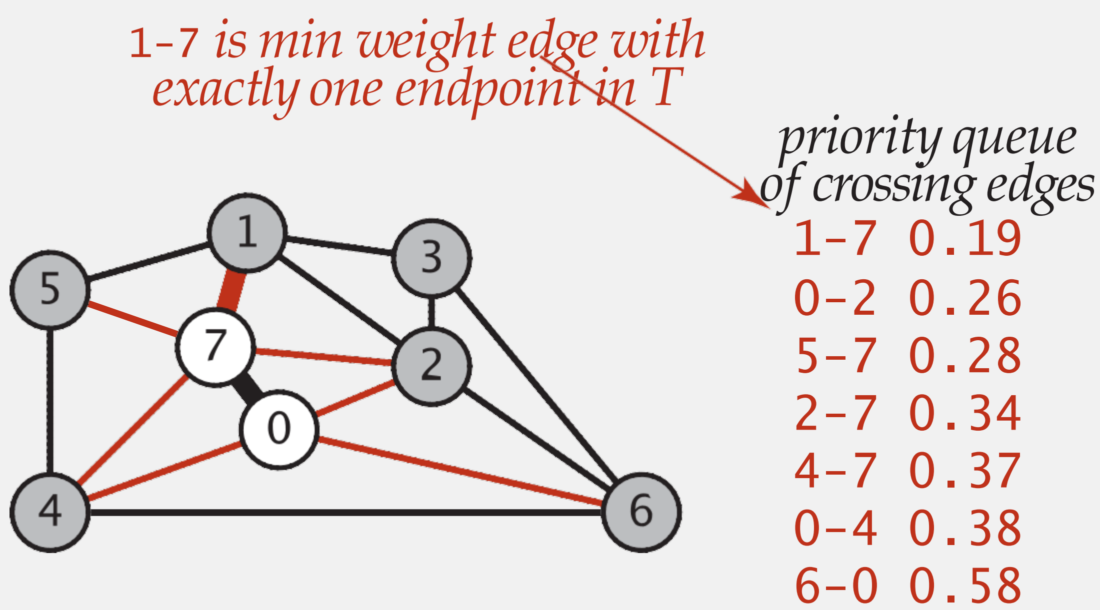
1-7	0.19
0-2	0.26
5-7	0.28
2-7	0.34
4-7	0.37
0-4	0.38
6-0	0.58

Prim's algorithm: lazy implementation

Challenge. Find the min weight edge with exactly one endpoint in T.

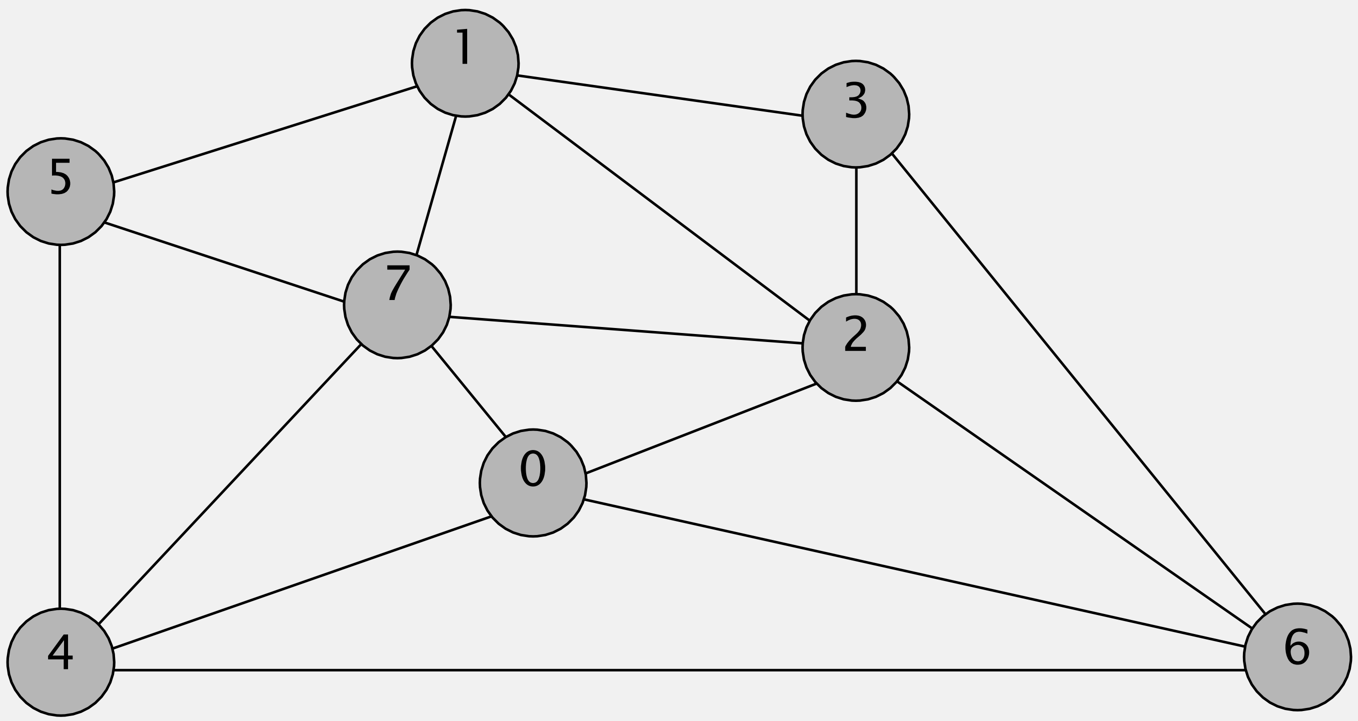
Lazy solution. Maintain a PQ of edges with (at least) one endpoint in T.

- Key = edge; priority = weight of edge.
- Delete-min to determine next edge $e = v–w$ to add to T.
- Disregard if both endpoints v and w are marked (both in T).
- Otherwise, let w be the unmarked vertex (not in T):
  - add to PQ any edge incident to w (assuming other endpoint not in T)
  - add e to T and mark w



Prim's algorithm (lazy) demo

- Start with vertex 0 and greedily grow tree T.
- Add to T the min weight edge with exactly one endpoint in T.
- Repeat until $V - 1$ edges.

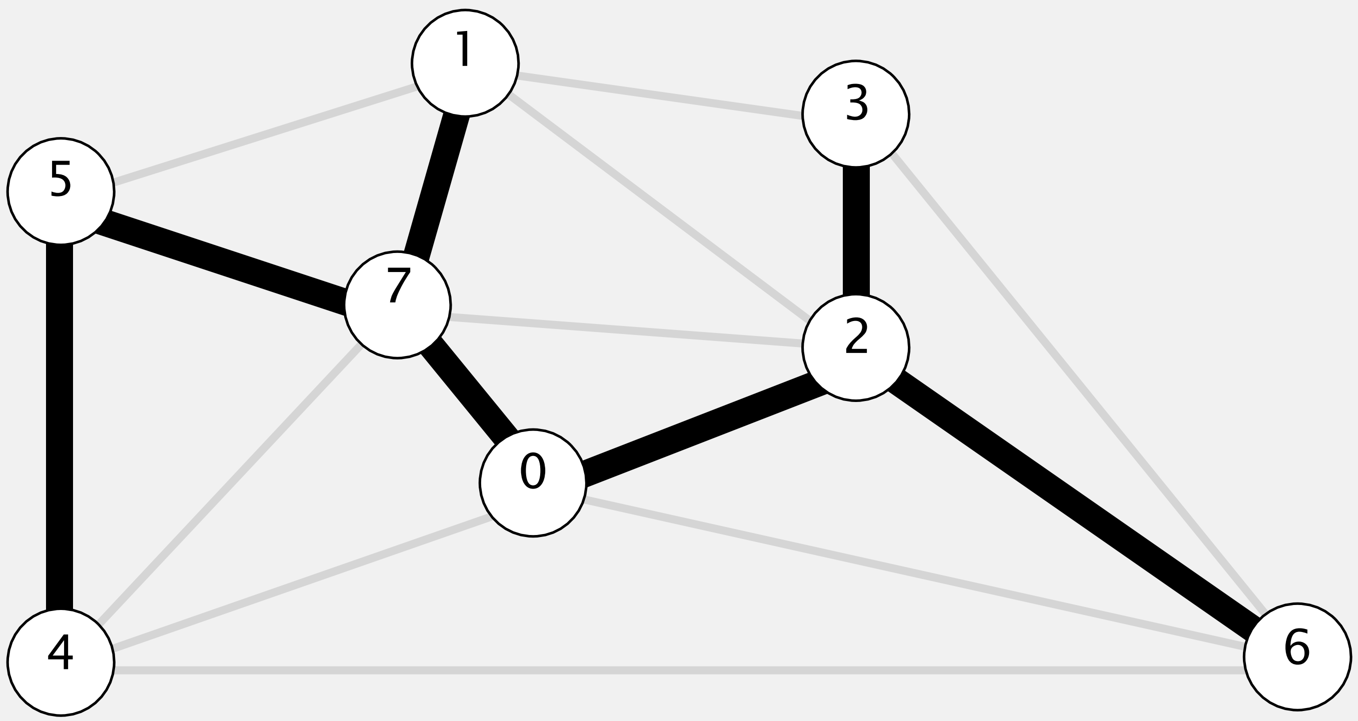


an edge-weighted graph

0-7	0.16
2-3	0.17
1-7	0.19
0-2	0.26
5-7	0.28
1-3	0.29
1-5	0.32
2-7	0.34
4-5	0.35
1-2	0.36
4-7	0.37
0-4	0.38
6-2	0.40
3-6	0.52
6-0	0.58
6-4	0.93

Prim's algorithm (lazy) demo

- Start with vertex 0 and greedily grow tree T.
- Add to T the min weight edge with exactly one endpoint in T.
- Repeat until $V - 1$ edges.



MST edges

```
0-7   1-7   0-2   2-3   5-7   4-5   6-2
```

Prim's algorithm: lazy implementation

```
public class LazyPrimMST
{
   private boolean[] marked;   // MST vertices
   private Queue<Edge> mst;    // MST edges
   private MinPQ<Edge> pq;     // PQ of edges

    public LazyPrimMST(WeightedGraph G)
    {
        pq = new MinPQ<Edge>();
        mst = new Queue<Edge>();
        marked = new boolean[G.V()];
        visit(G, 0);

        while (!pq.isEmpty() && mst.size() < G.V() - 1)
        {
           Edge e = pq.delMin();
           int v = e.either(), w = e.other(v);
           if (marked[v] && marked[w]) continue;
           mst.enqueue(e);
           if (!marked[v]) visit(G, v);
           if (!marked[w]) visit(G, w);
        }
   }
}
```

assume G is connected

repeatedly delete the min weight edge e = v–w from PQ

ignore if both endpoints in T

add edge e to tree

add v or w to tree

Prim's algorithm: lazy implementation

```
private void visit(WeightedGraph G, int v)
{
   marked[v] = true;
   for (Edge e : G.adj(v))
      if (!marked[e.other(v)])
         pq.insert(e);
}

public Iterable<Edge> mst()
{  return mst;  }
```

add v to T

for each edge e = v–w, add to PQ if w not already in T

Lazy Prim's algorithm: running time

Proposition. Lazy Prim's algorithm computes the MST in time proportional to $E \log E$ and extra space proportional to E (in the worst case).

Pf.

operation	frequency	binary heap
delete min	E	$\log E$
insert	E	$\log E$

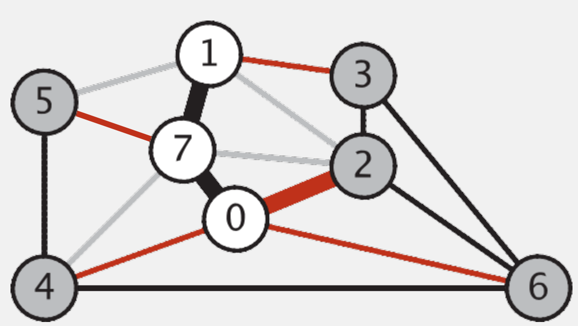
Prim's algorithm: eager implementation

Challenge. Find min weight edge with exactly one endpoint in T.

pq has at most one entry per vertex

Eager solution. Maintain a PQ of vertices connected by an edge to T, where priority of vertex v = weight of shortest edge connecting v to T.

- Delete min vertex v and add its associated edge $e = v–w$ to T.
- Update PQ by considering all edges $e = v–x$ incident to v
  - ignore if x is already in T
  - add x to PQ if not already on it
  - decrease priority of x if $v–x$ becomes shortest edge connecting x to T



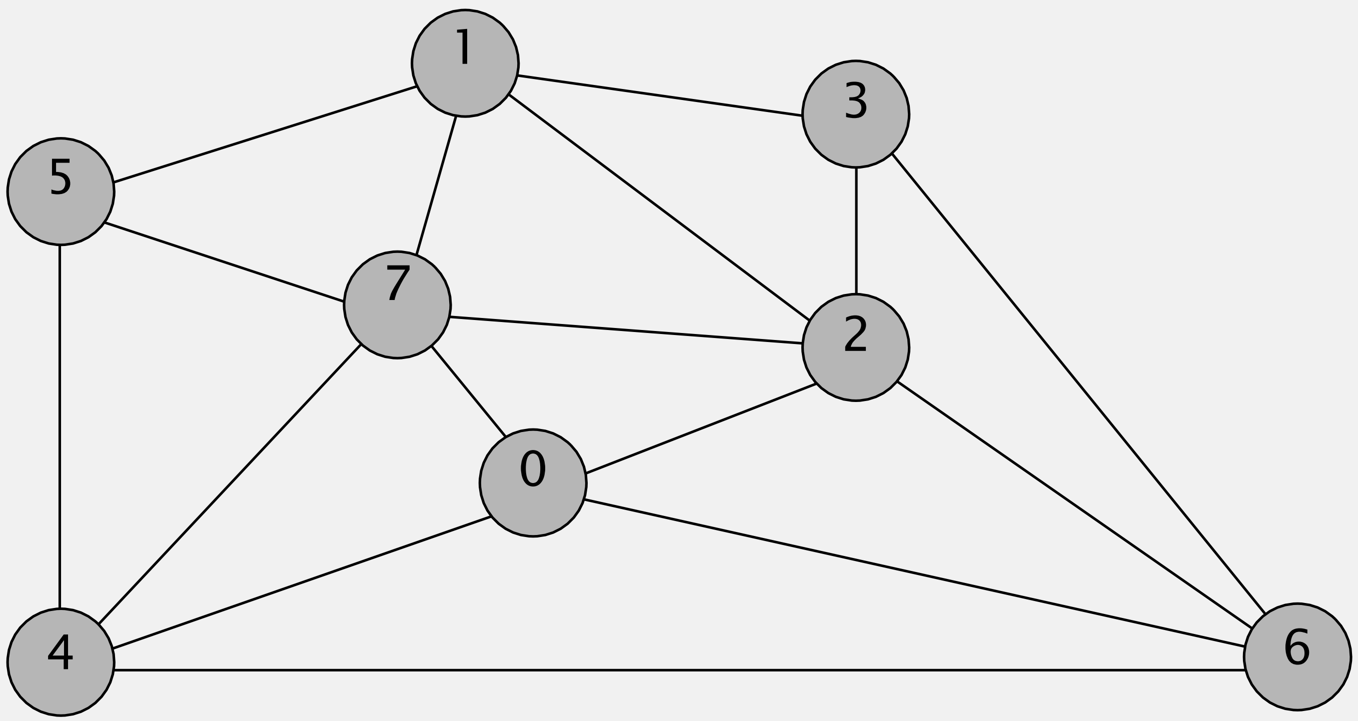
0		
1	1-7	0.19
2	0-2	0.26
3	1-3	0.29
4	0-4	0.38
5	5-7	0.28
6	6-0	0.58
7	0-7	0.16

red: *on PQ*

black: *on MST*

Prim's algorithm (eager) demo

- Start with vertex 0 and greedily grow tree T.
- Add to T the min weight edge with exactly one endpoint in T.
- Repeat until $V - 1$ edges.

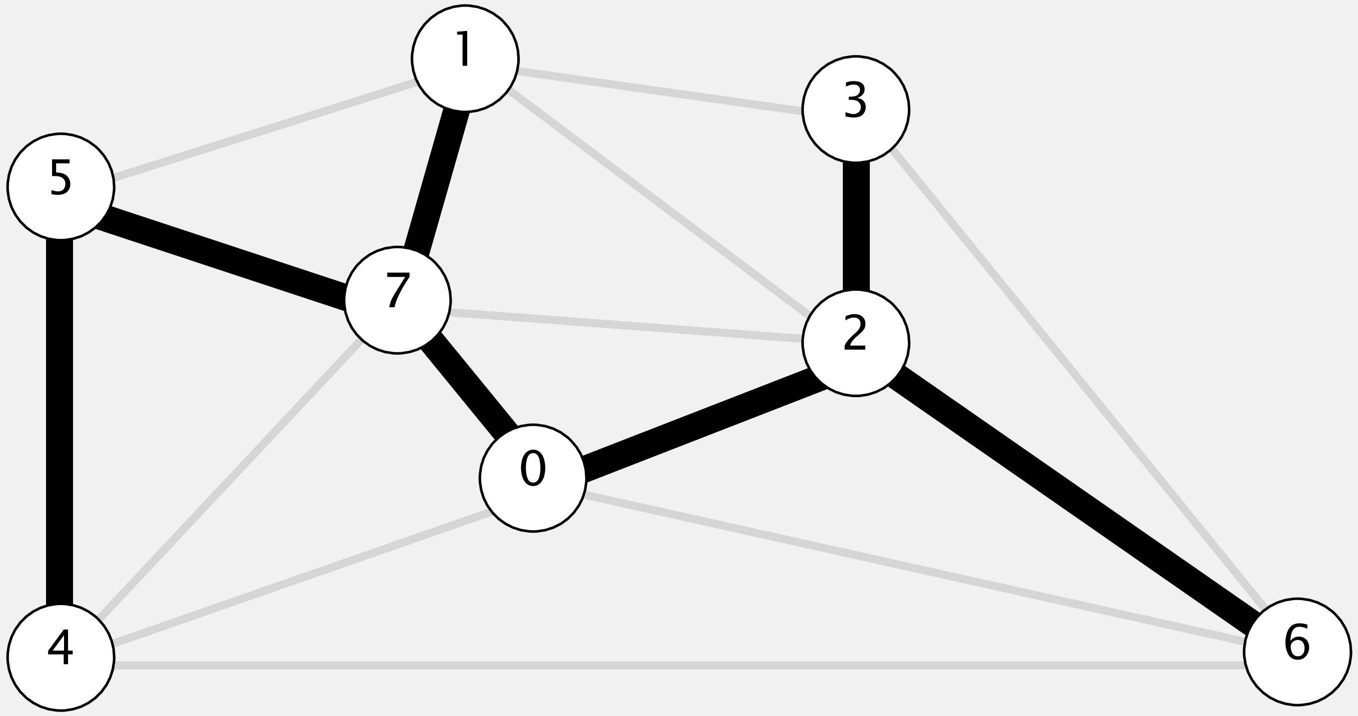


an edge-weighted graph

0-7	0.16
2-3	0.17
1-7	0.19
0-2	0.26
5-7	0.28
1-3	0.29
1-5	0.32
2-7	0.34
4-5	0.35
1-2	0.36
4-7	0.37
0-4	0.38
6-2	0.40
3-6	0.52
6-0	0.58
6-4	0.93

Prim's algorithm (eager) demo

- Start with vertex 0 and greedily grow tree T.
- Add to T the min weight edge with exactly one endpoint in T.
- Repeat until $V - 1$ edges.



v	edgeTo[]	distTo[]
0	-	-
7	0–7	0.16
1	1–7	0.19
2	0–2	0.26
3	2–3	0.17
5	5–7	0.28
4	4–5	0.35
6	6–2	0.40

MST edges

`0-7   1-7   0-2   2-3   5-7   4-5   6-2`

Indexed priority queue

Associate an index between 0 and $N - 1$ with each key in a priority queue.

- Supports insert and delete-the-minimum.
- Supports decrease-key given the index of the key.

```
public class IndexMinPQ<Key extends Comparable<Key>>
```

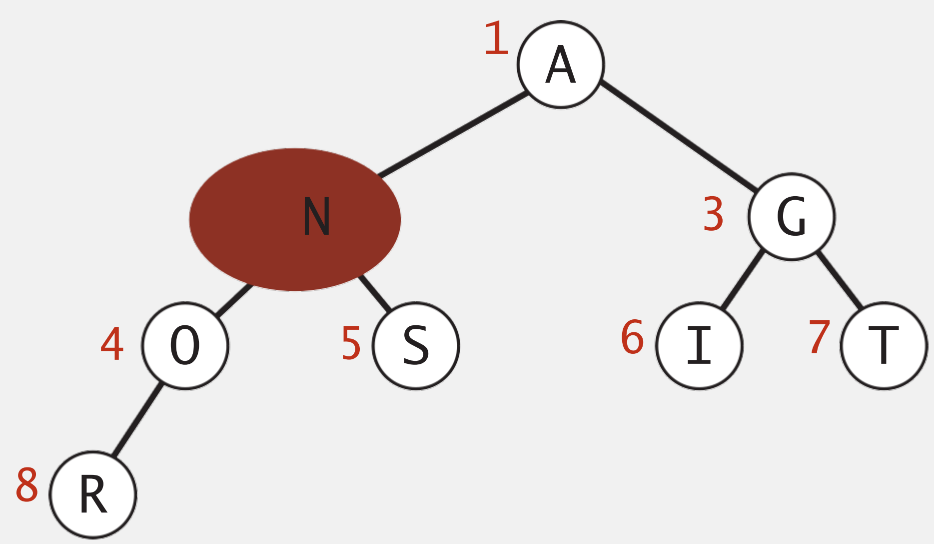
	`IndexMinPQ(int N)`	*create indexed priority queue with indices 0, 1, …, N – 1*
`void`	`insert(int i, Key key)`	*associate key with index i*
`void`	`decreaseKey(int i, Key key)`	*decrease the key associated with index i*
`boolean`	`contains(int i)`	*is i an index on the priority queue?*
`int`	`delMin()`	*remove a minimal key and return its associated index*
`boolean`	`isEmpty()`	*is the priority queue empty?*
`int`	`size()`	*number of keys in the priority queue*

Indexed priority queue implementation

Binary heap implementation. [see Section 2.4 of textbook]

- Start with same code as `MinPQ`.
- Maintain parallel arrays `keys[]`, `pq[]`, and `qp[]` so that:
  - `keys[i]` is the priority of `i`
  - `pq[i]` is the index of the key in heap position `i`
  - `qp[i]` is the heap position of the key with index `i`
- Use `swim(qp[i])` to implement `decreaseKey(i, key)`.

i	0	1	2	3	4	5	6	7	8
keys[i]	A	S	O	R	T	I	N	G	-
pq[i]	-	0	6	7	2	1	5	4	3
qp[i]	1	5	4	8	7	6	2	3	-



Prim's algorithm: which priority queue?

Depends on PQ implementation: V insert, V delete-min, E decrease-key.

PQ implementation	insert	delete-min	decrease-key	total
unordered array	1	V	1	V^2
binary heap	$\log V$	$\log V$	$\log V$	$E \log V$
d-way heap	$\log_d V$	$d \log_d V$	$\log_d V$	$E \log_{E/V} V$
Fibonacci heap	1 †	$\log V$ †	1 †	$E + V \log V$

† amortized

Bottom line.

- Array implementation optimal for dense graphs.
- Binary heap much faster for sparse graphs.
- 4-way heap worth the trouble in performance-critical situations.
- Fibonacci heap best in theory, but not worth implementing.